A Novel Approach to Equating English Teachers' and Chinese Teachers' Ratings of Behaviours Characterised by Attention Deficit Hyperactivity Disorder

LU, LIN

A Novel Approach to Equating English Teachers' and Chinese Teachers' Ratings of Behaviours Characterised by Attention Deficit Hyperactivity Disorder

Lin Lu

A thesis submitted for the degree of Doctor of Philosophy

School of Education

University of Durham

August 2023

Abstract

The diagnosis and treatment of ADHD rely on accurately identifying and interpreting symptoms. However, different raters may have different perceptions of ADHD symptoms, which can significantly impact ADHD diagnosis and prevalence rates. This study presented a novel way to compare ADHD symptom ratings between children from China and England while considering raters' differences.

The study developed a series of cartoon animations to measure the raters' leniency toward Children's ADHD symptoms. The Many-facet Rasch Model was then applied to adjust the children's ADHD symptom ratings according to their raters' leniency. The study was conducted in Year 2 classrooms in schools in China and England, and participating teachers were asked to rate cartoon characters' ADHD behaviours according to their tolerance. They were also asked to rate 10 children selected randomly from their class about ADHD symptoms.

The study found that Chinese teachers were more lenient with children's ADHD behaviours than their English colleagues. Moreover, after adjusting for raters' leniency, Chinese children's ratings increased significantly, while English children's ratings decreased significantly. The study also found that Chinese children's ratings of ADHD behaviours were significantly higher than those of English children. Additionally, the Inter-rater Agreement was low among Chinese teachers. The findings highlight the significant impact of raters' differences on ADHD symptom ratings and the importance of equating teachers' ratings of children's ADHD symptoms to produce a relatively fair comparison between countries. The study's use of cartoon animations offers many advantages over text and videotape vignettes for cross-cultural studies. Moreover, the findings suggest that multi-informants are necessary for a single setting for diagnosing ADHD in children.

In conclusion, this study provides valuable insights into the impact of rater differences on ADHD symptom ratings and the importance of considering these differences when comparing prevalence rates between countries. Future research should explore ways to improve inter-rater agreement among raters and investigate other factors that may affect ADHD diagnosis and treatment.

Declaration

I declare that this thesis is my own work. No material contained in this thesis has previously been submitted for a degree in this or any other institution.

Acknowledgement

Without the wisdom, patience, and kindness of my supervisors Dr. Christine Merrell, Pro. Peter Tymms and Dr. Xiaofei Qi, this work would not have been possible. I remain forever thankful for their support. Christine's passion for research has been inspirational and will live on in me. Peter has been a fond of knowledge and shown great stamina in supporting my arduous journey. Xiaofei has graciously shouldered the supervisory duties of the late Christine, making herself unstintingly available in support of this work.

My family and pets have also demonstrated great understanding and patience as I have laboured on this project.

Dedication

To Dr. Christine Merrell and my family.

Table of Contents

Table List

Figure List

Chapter 1 Introduction

1.1 Introduction

A substantial body of research suggests that Attention Deficit Hyperactivity Disorder (ADHD) has become the most common developmental or psychiatric diagnosis among school-age children (Polanczyk et al., 2007). According to worldwide prevalence estimates, 2.2% to 7.2% of children meet the diagnostic criteria for ADHD (Polanczyk et al., 2014; Thomas et al., 2015; Sayal et al., 2018). The core symptoms of ADHD, as defined by the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) (American Psychiatric Association, 2013), include poor attention span, excessive impulsivity, and hyperactivity. These symptoms significantly impact children's lives, their families, and societies.

Furthermore, the global diagnosis and research of ADHD have led to increased awareness, diagnosis rates, and medical interventions for the disorder. However, it is important to note that the prevalence of ADHD exhibits significant variation both across and within countries. This variability can be attributed to several factors, including disparities in diagnostic criteria, assessment methods, and research methodologies employed. Additionally, different countries have integrated distinct cultural and social factors into their approaches to diagnosing and treating ADHD. Consequently, the wide range of reported ADHD prevalence rates has sparked considerable debates concerning the validity of the disorder and the reliability of its diagnosis within diverse cultural contexts. In this chapter, I will explore the multifaceted impact of ADHD, address associated issues, and engage with the ongoing debates surrounding the global phenomenon of ADHD.

1.2 The Impact of ADHD

1.2.1 The Impact of ADHD on A Child's Life

First, it is more challenging for individuals with ADHD to adapt to the school environment compared to children without ADHD. Children with ADHD often experience related impairments, including underdeveloped motor skills, self-help abilities, personal responsibilities, and independence (Greene et al., 1996). Consequently, they may struggle to comprehend and adhere to basic school rules. Furthermore, they frequently face difficulties in controlling impulsivity, exhibit hyperactivity, struggle to remain seated, engage in excessive talking, disrupt other children's learning, and find it difficult to wait their turn. These behaviours not only affect their own learning but also impact the learning of their peers. Teachers must invest a significant amount of time and effort in managing their behaviour in the classroom, which can lead to elevated stress levels for educators (Sherman, Rasmussen & Baydala, 2008).

Second, many children with ADHD encounter challenges in their learning journey at school. School-based learning necessitates maintaining sustained attention during classwork and homework. However, due to difficulties in sustaining attention, these children often find it arduous to complete assigned tasks, particularly when those tasks involve complex information, require organizational skills, or deliberate rehearsal strategies (Douglas & Benezra, 1990). Additionally, children with ADHD are more susceptible to experiencing reading disabilities and mathematical difficulties than their peers without ADHD (Frick et al., 1991). As a result of these learning difficulties, 30% of children with ADHD are considered academic underachievers in comparison to their peers of the same age and IQ without ADHD (Frick et al., 1991; Kamphaus & Frick, 1996).

Third, establishing friendships can be a challenge for children with ADHD. Their struggles stem from impulsive behaviour and a short attention span, making it difficult to learn and adhere to social norms while comprehending social cues (Wehmeier et al., 2010). Impulsive and interruptive behaviour, coupled with a tendency to abandon play activities due to boredom, pose challenges for children with ADHD in their interactions with peers. Additionally, maintaining friendships becomes problematic as these children may be perceived as 'different' by their schoolmates (Anastopoulos & Shelton, 2001).

Fourth, children with ADHD face an elevated risk of experiencing other emotional issues such as low self-esteem, mood disorders, depression, and anxiety disorders (August et al., 1996; Biederman et al., 1996). These emotional challenges often arise from difficulties in learning, adhering to school rules, and achieving social competence. Consequently, children with ADHD may have fewer opportunities for success and encounter more negative feedback than their peers. As competence serves as a foundation for healthy emotional development, these children might perceive themselves as less competent, subsequently affecting their emotional well-being. Research indicates that between 13% and 51% of children with ADHD exhibit emotional disorders (Jensen et al., 1997). Given the range of difficulties faced by children with ADHD, the impact on their educational functioning is substantial. Up to 46% of children with ADHD experience school suspensions, 10% to 20% face school expulsions, and 10% to 35% do not graduate from high school (Barkley, 1998). Furthermore, children with ADHD are more susceptible to accidental injuries compared to their peers without ADHD, with up to 57% described as accident-prone and 15% experiencing serious accidental injuries (Reebye, 1997).

1.2.2 The Impacts of ADHD to Families and Societies

As discussed above, ADHD has a significant impact on a child's life. While it was once believed that children with ADHD would outgrow the disorder, research indicates that the

effects of ADHD symptoms can persist into adulthood for a substantial portion of individuals—30% to 60% continue to display symptoms and related impairments (Resnick, 2005). Studies highlight that young people with ADHD face an elevated risk of academic failure, school dropout, teenage pregnancy, delinquency, and risky driving behaviours (Barkley, 2002; Wehmeier et al., 2010). Additionally, adults with ADHD may encounter challenges in maintaining employment and managing relationships with employers, colleagues, and family members (Torgersen et al., 2006). These adults are also at an increased risk of drug and substance abuse (Biederman et al., 1998).

This lifelong disorder not only affects the individual's life but also affects their families and society at large. Raising a child with ADHD can lead to substantial stress for parents and impact marital relationships (Pelham & Lange, 1993). The financial burden of treatment can be significant for families and their members, with expenditures reaching $6.78 billion for children and $12.10 billion for adults with ADHD in the US (Birnbaum et al., 2005). Furthermore, family members of individuals affected by ADHD tend to have 1.6 times as many medical claims compared to matched control individuals without a family member diagnosed with ADHD (matched based on age, gender, geographical location, and employment status) (Swensen et al., 2003). Adults with ADHD are more likely to experience marital issues, separation, divorce, or report difficulties in their parent-child relationships (Murphy & Barkley, 1996), thus significantly impacting their families.

Studies suggest that this disorder results in increased costs across education, healthcare, and other domains. Schools are required to allocate considerable effort and resources to support children with ADHD, encompassing special education services, strategies to address disruptive classroom behaviours, and in-school medication administration (Altemeier & Horwitz, 1997).

Meanwhile, the financial burden extends to medication treatment and therapy for the disorder, impacting national healthcare systems. Medical expenses encompass hospitalizations, primary care office visits, outpatient mental health appointments, and prescription medications.

While estimating the average cost of ADHD treatment and intervention is challenging due to international variations in medical healthcare systems and practices, studies conducted across numerous countries indicate that treating ADHD is expensive. Medical cost studies suggest that, compared to children without ADHD, those with ADHD incur annual costs ranging from $503 to $1,343 (matched controls) and $207 to $1,560 (unmatched controls) (S Matza et al., 2005). In the US, the combined annual costs for healthcare and educational services for ADHD children and adolescents are estimated to be $143-266 billion. Meanwhile, in the UK, as of 2010, the projected total cost to the National Health Service and education services for ADHD children and adolescents is approximately £670 million (Holden et al., 2013). In conclusion, ADHD not only affects an individual's quality of life but also leaves a substantial impact on their families and society as a whole.

1.3 The Globalisation Trend of ADHD

1.3.1 The Trend and Drivers

Alongside the increasing awareness of this disorder and its related impacts, signs of the globalisation of ADHD are becoming apparent. Just three decades ago, ADHD was referred to as the "American Disorder," with the majority of research, diagnoses, and treatments centred in the United States (Faraone et al., 2003). With the establishment of standardised diagnostic criteria (DSM and ICD), studies on ADHD prevalence have revealed its diagnosis in numerous countries (Polanczyk et al., 2007; Thomas et al., 2015). A global trend of rising prevalence and medication usage for ADHD has emerged. On a worldwide scale, the estimated prevalence of ADHD was 5.29% in 2007; by 2015, this figure had risen to 7.2%, encompassing more

countries and studies (Polanczyk et al., 2007; Thomas et al., 2015). Meanwhile, a study found that the utilisation of medication for treating children with ADHD has steadily increased across all investigated countries (13), with an annual absolute increase ranging from 0.02% to 0.26% (Raman et al., 2018). Furthermore, the number of countries employing medication for treating children with ADHD has surged from 31 to 55 over a decade (Scheffler et al., 2007).

Several factors driving the globalisation of this disorder have been examined. Hinshaw and Scheffler (2014) propose that the growth of compulsory schooling and the emphasis on academic achievement may contribute to the increasing global diagnosis and treatment rates. Meanwhile, Conrad and Bergey (2014) suggest that the international pharmaceutical industry, the expanding influence of American psychiatry, the adoption of DSM criteria for ADHD diagnosis, the availability of information on the Internet, screening checklists, and ADHD advocacy groups all play roles as potential drivers for the global spread of ADHD diagnosis. This globalisation has resulted in heightened awareness of ADHD among parents, teachers, and societies worldwide, accompanied by an upsurge in ADHD diagnoses and the usage of medications for treating affected children globally. The globalisation of this disorder suggests that ADHD is a genuine medical condition rather than solely a cultural phenomenon. Nevertheless, global research has also prompted numerous questions regarding ADHD.

1.3.2 Different Prevalence and Diagnostic Criteria

The prevalence of ADHD varies both across and within countries, as shown in *Table 1* below. From the table, it is evident that there is variation in prevalence rates across different countries. One notable factor contributing to this disparity is the adoption of distinct diagnostic criteria by different countries. Two primary diagnostic criteria are widely utilised: The Diagnostic and Statistical Manual of Mental Disorders (DSM) and The International Classification of Diseases (ICD). The DSM is published by the American Psychiatric Association (APA) and is

extensively employed in North America. On the other hand, the ICD, established by the World Health Organization (WHO), is prevalent in Europe.

While both sets of criteria present a similar symptom description for ADHD, there exists a divergence in the threshold levels. Specifically, the ICD imposes a higher threshold compared to the DSM. The ICD mandates a manifestation of symptoms across all dimensions, whereas the DSM necessitates symptoms to be evident in at least one dimension. Further discrepancies between the two diagnostic criteria encompass factors such as the definition of the relationship between primary diagnosis and associated comorbidities, as well as variations in the applied age ranges (ICD-10, 2015; DSM-5, 2013). These nuanced dissimilarities in diagnostic criteria significantly contribute to the observed variability in ADHD prevalence rates across different countries. Notably, research indicates that prevalence rates diagnosed using DSM criteria are slightly higher than those identified using ICD criteria (Lee et al., 2008).

Table 1 The prevalence of ADHD worldwide

Countries	Prevalence	Diagnosis Criteria	Reference
US	11%	DSM-IV	(Visser et al., 2014)
Canada	5.4% - 6.1%	DSM-IV	(Charach, Lin & To, 2010)
Australia	6% - 9%	DSM-IV	(AI-Yagon et al., 2013)
Germany	4.8%	ICD-10	(Schlack et al., 2007)
UK	1.9% - 5%	ICD DSM	(Murphy et al., 2014)
Portugal	4%-5%; 5%-8%	DSM	(Rodrigues & Nuno, 2014), (Neto, 2014)

Ireland	5%-10%, 3.7%	Not clear	(Joint Committee on Health and Children, 1999), (Health Service Executive,2013)
Japan	3% - 7%	DSM	(MHLW, 2015)
Taiwan	6.3% - 12%	DSM-IV	(Huang, 2008)
Chile	6.2% - 10%	DSM	(Ministerio de Salud, 2008) (De la Barra et al., 2013)
Italy	1%-2%, 3.51%	DSM	(Maschietto, Re & Cornoldi, 2012) (Reale et al., 2014)
France	3.5%-5.6%	DSM	(Lecendreux et al., 2011)
China	5.6% - 6.26%	DSM	(Lian et al., 2013; Wang et al., 2017)

1.3.3 Different Methods and Methodologies

However, distinct criteria are not the sole factor contributing to the observed variability. The methods and methodologies adopted by studies also influence the reported prevalence of ADHD. These encompass sampling methods, informants, instruments, and the utilisation of full or partial criteria. The choice of sampling method can significantly impact prevalence estimates. Certain studies opt for clinical samples, while others select community samples. Additionally, the age range of the chosen samples can influence the observed prevalence, given that certain ADHD symptoms tend to diminish with age (Faraone et al., 2005).

The involvement of different informants (parents, teachers, or clinical professionals) in the diagnostic process can also influence prevalence estimates. Some studies rely on input from parents and teachers, while others involve clinical professionals or a combination thereof. It is worth noting that parents and teachers often tend to overestimate the severity of symptoms,

leading to potential overdiagnosis (Getahun et al., 2013). Furthermore, agreement between parents and teachers tends to be inconsistent. In Japan, for instance, the estimated ADHD prevalence based on parents' ratings was 31.1%, whereas the prevalence derived from teachers' ratings dropped to 4.3% (Soma et al., 2009). Such disparities highlight significant differences in perceptions of this disorder between parents and teachers. Consequently, various studies implement distinct approaches to address cases where informants do not concur, potentially contributing to variations in prevalence.

Moreover, the selection of diagnostic instruments holds significant weight in the overall procedure. Different studies employ varied instruments for diagnosing symptoms. While some studies utilise interviews, others rely on questionnaires and observation. Additionally, as diagnostic criteria are established and published within an English language and cultural context, other regions must translate and adapt these criteria to their own linguistic and cultural milieu. Unfortunately, there exists no universal guideline, process, or assessment to ascertain the reliability and validity of adapted criteria for the specific populations they target. Consequently, the diverse methods and methodologies employed by different studies result in varying estimates of ADHD prevalence. Consequently, some researchers have argued that comparing absolute ADHD prevalence rates across different cultural contexts may lack validity (Singh et al., 2013; Filipe, 2016).

1.3.4 Different Social and Cultural Factors

Reports of ADHD prevalence worldwide yield diverse findings, stemming not only from variations in diagnostic methods and methodologies but also from social and cultural factors. Studies indicate differing rates of ADHD diagnosis among racial and ethnic groups. Notably, white children exhibit the highest rate, followed by black children and other racial groups (Visser et al., 2014; Morgant et al., 2013). Bloom et al. (2013) discovered that Hispanic

children have a lower diagnosis rate (6%) than their non-Hispanic white counterparts (12%). Furthermore, they noted that children from single-parent households and less-educated families have a higher ADHD diagnosis rate compared to those from two-parent households and families with higher education levels. Visser et al. (2014) also observed that children with healthcare coverage experience a higher ADHD diagnosis rate than those without coverage. A UK-based study highlighted the correlation between social class and significant differences in the prevalence of childhood mental disorders, indicating that 14% of children from unskilled occupation families receive a mental disorder diagnosis, in contrast to only 5% of children from professional families (Meltzer et al., 2000). This suggests that various prevalence rates may result from social factors, such as distinct awareness levels of the disorder and differing barriers to obtaining a diagnosis.

Furthermore, cultural distinctions among raters may also influence the diagnosis of ADHD. Schmitz et al. (2003) compared the ratings provided by Mexican, Mexican American, and Puerto Rican mothers regarding their children's ADHD behaviours. Their findings indicated that these three groups of mothers significantly differed in their ratings of hyperactive and impulsive behaviours, though not inattentive behaviours. This highlights the varying perceptions of ADHD symptoms among mothers from different cultural backgrounds. Additionally, as ADHD diagnosis becomes more prevalent worldwide, gender differences in the disorder shift across cultures. Historically, ADHD has been regarded as predominantly affecting boys, with gender ratios consistently reported at 3:1 or 4:1 (Arnold, 1996; MTA Cooperative Group, 1999). However, Chile's largest national survey discovered a gender ratio of boys to girls of 1:3 among adolescents aged 12 to 18 years (De la Barra et al., 2013). Similarly, studies conducted in Israel suggest a significant increase in female prevalence within the 13 to 17-year age group, accompanied by a decrease in the gender ratio from 2.94 to 1.86

(Michael et al., 2017). While the reasons behind the rising identification of girls with ADHD remain unclear, these findings suggest the presence of referral bias and challenges in identifying cases. Hyperactive and impulsive symptoms in boys are more conspicuous, making their behaviour more noticeable. Conversely, girls are more prone to displaying inattentive symptoms linked to cognitive impairments, which are less easily detected (Sciutto et al., 2004).

1.3.5 Different Treatments

Worldwide research has revealed that different countries prioritize various treatments for ADHD. In certain nations, strong criticisms of biological psychiatry restrict the use of medication-based interventions. Some countries, including the United States, Canada, Australia, Germany, and Chile, emphasize medicating children with ADHD. Conversely, others focus on child therapy and intervention, such as the United Kingdom, Japan, France, and Brazil (Bergey et al., 2018). Diverse emphases on treatment methods for psychiatric disorders are evident across different countries.

In the United States, medicating children for disorders is widely accepted. The American market accounts for 85% of global ADHD drug consumption, with one in 20 to 25 children in the U.S. utilising stimulant medication for treatment (Zuvekas et al., 2006). While medication serves as the primary treatment approach for American children with ADHD, national disability policies also provide for educational accommodations, offering non-medical treatment alternatives.

In the United Kingdom, medications are reserved for treating severe cases of ADHD. The UK exhibits the lowest rate of medication treatment for children with ADHD (Murphy et al., 2014). A mere 0.02% to 1.3% of children with ADHD receive medication-based treatment in the UK (Beau-Ledjstrom et al., 2016). Although children with ADHD are not officially recognised as

having a disability and do not automatically receive special education services, parents can seek special education support, such as parent training and psychological intervention, on behalf of their children. The UK expresses broader concerns within medical circles, among parents, and across society regarding medication-based solutions for children's mental disorders, potentially influencing the low medication rate.

Moreover, traditional beliefs and religions may influence clinical and parental treatment decisions for children with ADHD. In Japan, the rate of medication for children with mental disorders remains low (Scheffler et al., 2007). While Japanese clinicians have begun to accept ADHD medication, a tradition of favouring environmental coordination, parent training, and psychosocial interventions persists (Mari, 2018). Japanese parents often harbour limited expectations of medication as a solution, with only 23% considering it a viable option for their children's treatment (Saito, 2005). In India, parents tend to seek religious intervention for treatment and may resist the biomedical explanation of their children's behaviour (Wilcox, Washburn & Patel, 2007).

Similarly, in mainland China, ADHD medication is not covered by national health care, while the use of Chinese traditional herbs in ADHD treatment is included. China's long history of using traditional herbs for medical purposes persists, even in the context of ADHD, due to national healthcare policies and cultural factors (Aramwit & Wirotsaengthong, 2013). In Ghana, beliefs in spiritual forces, curses, and charms contributing to mental illnesses lead most individuals to consult herbalists, traditional healers, or Christian groups for treatment (Ae-Ngibise et al., 2010; WHO, 2007).

As discussed, the treatment landscape for children with ADHD is influenced by multiple factors. Firstly, the national healthcare system plays a pivotal role, offering different options for diagnosis and treatment in various countries, thereby significantly impacting access to diagnosis and treatment within a country. Secondly, the types of non-medical treatment available to families of children with ADHD vary across nations, often involving collaboration between parents, teachers, schools, and educational experts. This complexity makes non-medical treatments costlier and more intricate to administer compared to medication. The availability of such treatments to families and their accessibility depends on public funding and social policies specific to each country.

Thirdly, the presence of a stigma associated with mental disorders can affect families' willingness to accept the diagnosis and treatment for children with ADHD. While attitudes toward the disorder differ globally, Asian countries such as Japan, Taiwan, and Korea exhibit stronger stigmas than their European counterparts (Bergey et al., 2018). However, these Asian nations share a competitive, evaluation-driven education system, where parents and teachers hold high expectations for children's behaviour, often leading to a pattern of seeking diagnosis and treatment for ADHD behaviours.

For instance, in Japan, 31% of parents perceive their children's behaviours as meeting ADHD diagnostic criteria (Soma et al., 2009). In Taiwan, about 97% of teachers hold a positive view of medication treatment for ADHD symptoms (Meng & Hsieh, 2012). Nonetheless, there exists a pronounced stigma and shame surrounding mental illness within Asian cultures (Mikyong Kim-Goh et al., 2015). The fear of social stigmatisation and the shame associated with displaying mental illness symptoms have resulted in delayed diagnosis and treatment seeking, as well as a reluctance to disclose diagnostic results. Thus, recognising the influence of social,

familial, and self-stigma across different contexts proves essential, as these factors shape perceptions of ADHD diagnosis and treatment.

1.4 Debates about ADHD in The Globalisation Trend

The globalisation of ADHD and the estimated worldwide prevalence do not eliminate the arguments about diagnosing and treating children with ADHD. The global variation in ADHD prevalence has fuelled criticism about the validity of this disorder and the diagnostic criteria in various cultural settings.

In Brazil, there is a higher estimated prevalence (18%) than the prevalence of 15% of a similar study carried out in America. (Guardiola et al., 2000; Nola et al., 2001). They found that the American diagnostic criteria are inaccurate for diagnostic use because it overestimates the disorder in Brazil. They suggested that the American diagnostic criteria focus on observed symptoms rather than related impairments, duration and severity, which might cause the overestimation.

In Ghana, ADHD symptoms are associated with disrespectful and disobedient behaviour of children towards adults, such as parents and teachers. As respect and reciprocity are the key values of their culture, children who do not listen, speak loudly to adults, and do not achieve expected academic achievements are perceived as having ADHD by parents and teachers (Christian et al., 2018).

Therefore, people argue that a high variation in prevalence across cultures and other disparate findings from different regions indicate that this disorder is a culturally and socially constructed disorder rather than a universal mental or psychological disorder (Timmi & Taylor, 2004). They argued that this disorder is a product of a Western culture driven by the profit-seeking

pharmaceutical industry, the high status of psychiatrists looking for new roles, and the Western culture of raising best-behaved children (Anderson, 1996). A mathematically based calculation of global prevalence does not reject the hypothesis that ADHD is a societal construct (Parens & Johnston, 2011; Singh et al., 2013). Polanczyk et al. have been criticised for accepting funds from a pharmaceutical company, 'COMPANY', that makes the ADHD drug Strattera, and several authors of this publication were previously employees of this company (Smith, 2017). It raised concerns about the objectivity and credibility of the study, as financial ties to the industry may influence the authors' interpretation of the results.

However, all disorders are, to some extent, constructed. Therefore, it is not valid to argue that because a disorder is, to some extent, constructed, it is not "real" (Peter & Diane, 1975). Meanwhile, research about the impairments of ADHD has been carried out all over the world. The findings indicate that a cluster of behaviours (hyperactivity, impulsivity, inattention) negatively impacts children's quality of life (Wehmeier et al., 2010; Barkley, 2013; Bauermeister et al., 2007). These related impairments are real. Therefore, the fact that ADHD is constructed to some extent does not negate the reality of the related impairments observed and documented through research. It suggests that ADHD is a real condition that can cause significant functional impairments and deserves attention and treatment. While it is true that the diagnostic criteria for ADHD have changed over time and may reflect societal norms and values, this does not mean that the condition itself is not real or that its impact on individuals' lives is not significant. As with many medical conditions, the diagnostic criteria and classification of ADHD may evolve as our understanding of the disorder improves and as new information emerges. However, the reality of the impairments and challenges faced by individuals with ADHD should not be dismissed or minimised.

While it's important to acknowledge that Polanczyk et al.'s study received funding from a pharmaceutical company, it's crucial to note that financial ties to the pharmaceutical industry are commonplace in medical research and don't necessarily indicate scientific misconduct or intentional bias. Many researchers and institutions have established policies and guidelines to manage and disclose potential conflicts of interest. Peer review processes are designed to assess the quality and validity of research findings irrespective of funding sources. Their study concluded that the global variation in ADHD prevalence primarily stems from methodological differences rather than cultural and social disparities (Polanczyk et al., 2007). This worldwide prevalence bolsters the argument that ADHD is not solely a social or cultural construct but rather a universal disorder (Moffitt & Melchior, 2007).

Furthermore, Polanczyk et al. emphasized that cultural and social factors should not be dismissed in the diagnosis of ADHD. Given the absence of a biological marker for this disorder and the diagnosis relying on symptoms, the identification of a child with ADHD hinges on the disparity between the child's behaviour and the societal or cultural expectations for children's behaviour. Consequently, social and cultural factors could influence judgments regarding the severity of the symptoms. These cultural and social factors continue to exert an impact on the universal nature of the disorder.

With the globalization of ADHD and growing international interest in behaviours linked to ADHD, numerous researchers contend that ADHD is a multifaceted condition encompassing interwoven biological and environmental dimensions (Sigh, 2008; Sonuga-Barke et al., 2010). In light of diverse studies spanning the globe, a recently published book titled "Global Perspectives on ADHD" (Bergey et al., 2018) delves into the social dimensions of diagnosing and treating ADHD across sixteen countries. Within this context, the authors propose that

ADHD is best perceived as a biosocial assemblage comprising a fluid, heterogeneous elements interconnected yet lacking a fixed ontology (Bergey et.al, 2018, p. 327). They further suggest that future research should centre on identifying these factors and delineating their interactions within specific contexts to enhance our comprehension of ADHD and its accurate diagnosis.

1.5 Conclusion

In order to identify these factors and describe their relations, comparative research is a powerful tool for investigating the aetiology of the disorder and generating new hypotheses about aetiology when populations vary in their exposure to risk factors. Higher prevalence populations could offer clues to risk factors, whereas low-prevalence populations could shed light on protective factors. However, methodological biases and other potential confounding factors would need to be eliminated before drawing such conclusions. Methodological artefacts are relatively easy to cope with for comparative research because research can be conducted in comparative groups with the same method and instrument.

While it is hard to rule out other factors, such as the variation caused by informants. The diagnosis of ADHD depends upon observers' interpretation, and the symptoms and impairments are dimensional, not categorical. The interpretation, expectations and tolerances of individual informants affect their judgement about the severity of the symptoms. Therefore, individual differences in perception and interpretation of symptoms can lead to variability in diagnosis.

Comparative studies often encounter challenges due to uncertainty regarding the observed differences in ADHD prevalence between two subject populations. It is difficult to ascertain whether these differences truly reflect disparities between the populations or if they are influenced by rater bias. Informants' judgments are inevitably shaped by their own social and

cultural factors. Thus, it is essential to develop an approach that can account for individual variations in rater leniency and appropriately adjust their ratings. By considering the individual factors that influence observers' judgments, this approach has the potential to enhance the accuracy and reliability of ADHD diagnosis and research while facilitating a fairer comparison between different cultural groups.

In order to achieve this goal, this study conducted a comparative investigation involving two culturally distinct countries: China and the UK. The study employed an innovative method utilising cartoon animations to assess teachers' leniency towards ADHD symptoms. With the statistical models of modern test theory, Rasch and Many-facet Rasch models, the study developed assessment instruments capable of measuring teachers' leniency across diverse cultural groups and equating the ratings of children's ADHD symptoms accordingly. In this way, the study aimed to generate a relatively fair comparison between these culturally distinct groups.

Chapter 2 ADHD Diagnosis: A Journey Through History and Diagnosis Advancement

2.1 Introduction

Over the years, the understanding of ADHD and its diagnostic criteria has evolved, shaped by a complex interplay of historical, cultural, and scientific factors. Understanding the history of ADHD and the evolution of its diagnostic criteria is of paramount importance for several reasons. Firstly, exploring the historical perspectives allowed us to contextualise the disorder within different time periods, providing insights into how societal attitudes towards hyperactivity and attention difficulties have changed over time. This historical lens helped us appreciate the complexity of ADHD and recognize the social constructs and biases that had influenced its conceptualisation.

Secondly, tracing the diagnostic advancements helped the researcher understand the iterative process of refining diagnostic criteria, incorporating new research findings, and addressing limitations. By examining the diagnostic developments, the researcher gained a comprehensive understanding of the diagnostic frameworks currently in use and their underlying rationale. Moreover, understanding the historical and diagnostic journey of ADHD helped the researcher critically evaluate the strengths, weaknesses, and potential biases associated with current diagnostic practices. This knowledge facilitated ongoing efforts to improve diagnostic accuracy, enhance cultural sensitivity, and explore new avenues for understanding and treating ADHD. Ultimately, a journey through history and diagnosis advancement offered valuable insights that contributed to a more nuanced understanding of ADHD as a complex disorder.

This chapter is divided into two sections: "ADHD is Not a Modern Disorder" and "Diagnostic Developments and Criteria Evolution".

In the first section, "ADHD is Not a Modern Disorder," I explored the notion that ADHD is not a recent phenomenon but rather a condition that has existed throughout history. By examining historical accounts, medical literature, and cultural references, it shed light on early observations and descriptions of individuals exhibiting ADHD-like symptoms across different time periods and cultures. This section aims to challenge the misconception that ADHD is solely a product of modern society and highlights the importance of understanding its historical roots to gain a comprehensive perspective on the disorder.

The second section, "Diagnostic Developments and Criteria Evolution," delves into the evolution of diagnostic criteria and the advancements made in diagnosing ADHD. I traced the development of one of the widely used diagnostic systems, the Diagnostic and Statistical Manual of Mental Disorders (DSM) and explore the changes in diagnostic criteria over time. Additionally, I examined the shift from categorical diagnostic approaches to more dimensional frameworks, reflecting a better understanding of the heterogeneity and complexity of ADHD.

2.2 ADHD Is Not a Modern Disorder (1775-1902)

Although there is a wide range of impairments related to ADHD, as discussed above, some researchers still argue that ADHD does not exist. They believe that ADHD reflects social modernisation and compulsory education (Saul, 2014; Hinshaw & Scheffler, 2014). They argue that the modern environment requires children to stay indoors and focus on tasks for extended periods. Children who have difficulty adapting to this demanding and unfamiliar environment, which differs significantly from their homes, are labelled with ADHD and treated with drugs.

However, reviewing the history of ADHD reveals that it is not a modern disorder. ADHD symptoms can be found as characteristics in arts and literature from ancient times. In 1865, the German poet and children's book writer Heinrich Hoffmann wrote a poem called "Fidgety

Philip" (Taylor, 2011), portraying a boy with severe disruptive and hyperactive behavioural problems. In another story, "Johnny Look-in-the-Air," he described a boy with attention-sustaining issues in his book. Due to his vivid portrayal, "Fidgety Philip" has become a commonly used allegory for ADHD. In 1832, the German writer Johann Wolfgang von Goethe published his masterpiece, "Faust." In the play, there is a boy who exhibits a persistent pattern of excessive motor activities and impulsive behaviours without considering adverse consequences and warnings.

These characters in ancient works all display similar problematic behaviours, closely resembling the modern description of ADHD symptoms. While these dramatic behaviours undoubtedly captured the creators' attention and inspired their works, the descriptions are not clinical diagnoses or medical descriptions, and the problematic behaviours did not prompt medical attention for diagnosis and treatment. A recent discovery indicates that the earliest medical report of ADHD-like symptoms was by a German physician named Melchior Adam Weikard, preceding Still's report (1902) by more than a century (Barkley & Peters, 2012).

Although George Still's report is widely accepted as the first clinical report of ADHD symptoms, many clinical and medical records of these symptoms predate Still's work. While they did not garner serious attention at the time, their descriptions capture the core characteristics of ADHD symptoms. *Table 2* shows the clinical and medical records which are earlier than Still.

Table 2 The clinical and medica records which are earlier than Still

Name	Year	Term	Descriptions
Melchior Adam Weikard (German physician)	1775	Attention Deficit	● Lack of attention, perseverance and persistence; easily distracted, overactive and impulsive ● Poor upbringing, child rearing and biological predisposition ● More common among younger than among older individuals. ● Women were more inattentive than men. (Barkley & Peters, 2012)
Alexander Crichton (Scottish physician)	1798	Disease of attention	● Inattention, distractibility, impulsivity, restlessness, emotional reactivity ● Can be born with a person, dysregulated sensibility of the nerves, illness, head injury, a neurological disorder ● Symptoms diminished with age (Crichton, 1798)
Benjamin Rush (American physician)	1812	A syndrome involving the inability to focus attention	● "defective organisation in those parts of the body which are occupied by the moral faculties of the mind." (Rush, 1812, p339)
Charles West (English paediatrician)	1848	The nervous child	● A type of difficult children, neither an idiot nor insane (West, 1848)

Heinrich Neumann (German psychiatrist)	1859	Hypermetamorphosis	● Inability to stay focused, highly volatile in their inclinations, restlessness, unable to sit down and sit still (Neumann, 1859)
Desire-Magloire Bourneville (French physician)	1885	Mental instability	● Inattentiveness, disobedient, lacking discipline (Bader & Hadjikhani, 2014)
James William (American physician)	1890	Explosive will	● The effort of the attention is the essential phenomenon of will (James, 1890)
Thomas Smith Clouston (Scottish psychiatrist)	1899	Simple hyperexcitability	● Hyperexcitability, hypersensitivity, mental explosiveness ● Undue brain reactiveness to mental and emotional stimuli. Overactive to the nerve cells in the cerebral cortex (Clouston, 1899)
George F Still (English physician)	1902	The abnormal defect of moral control	● Aggressive, lawless, passionate, inattentive, impulsive, and overactive (Still, 1902)

Note. Martinez-Badía, J. (2015). Who says this is a modern disorder? The early history of attention deficit hyperactivity disorder. *World Journal of Psychiatry*, *5*(4), 379.

From this table, it becomes evident that ADHD is not a modern disorder. The terms and descriptions used by early medical professionals to characterise this type of behaviour, such as 'inattentive,' 'impulsive,' and 'overactive,' began to emerge in the 18th and 19th centuries. These

descriptions closely align with the modern description of ADHD symptoms. The recognition of this disorder dates back hundreds of years. However, it wasn't clinically documented as a behavioural disorder until 1902. English physician George Still (1902) delineated a group of 43 children from his clinical practice whom he categorised as having a deficit in "volitional inhibition" (p.1008) leading to "defects in moral control" (p.1009). To characterise the behaviours of these children, he employed terms such as aggressive, lawless, passionate, inattentive, impulsive, and overactive.

Furthermore, he identified three core impairments associated with these symptoms: "defect of cognitive relation to the environment; defect of moral consciousness; and defect in inhibitory volition" (p.1011). He presented his findings to the Royal Academy of Physicians in London. He postulated that these children lacked "moral consciousness," which governed a cognitive assessment of a person's actions in relation to actions that benefited "the good of all" (p.1008). Consequently, they struggled to inhibit behaviours that contravened moral principles. Still's report stands as the inaugural medical account of ADHD, and his elucidation of the mechanisms underlying children with ADHD became the cornerstone of contemporary ADHD theories (Sanghera, 2016).

Although he meticulously described the symptoms clinically, he remained uncertain about their cause. He hypothesised that these symptoms could stem from neurological deficiencies, which might be hereditary or the result of brain injury. Subsequently, other physicians, like Tredgold (1908) and Pasamanick et al. (1956), rallied behind Still's perspective that problematic behaviours might arise from undiagnosed brain damage. Similar to Still, they suggested that altering the environment and medication could ameliorate these behaviours, while also advocating for specialised education.

Since Still's seminal report, ADHD-like behaviours have captured the attention of psychiatrists, researchers, and educational practitioners. Subsequent research and efforts have been undertaken to scrutinise, define, and diagnose this disorder. It can be argued that ADHD is not a modern disorder. The contemporary environment and compulsory education have offered an opportunity to observe this disorder in a widespread cohort within similar settings worldwide.

2.3 Diagnostic Development and Criteria Evolution

In the realm of psychiatric disorders, Attention-Deficit/Hyperactivity Disorder (ADHD) has witnessed a dynamic evolution in its diagnostic criteria and conceptualisation over time. The development of diagnostic criteria for ADHD has been a multifaceted process shaped by the interplay of scientific research, clinical observations, and evolving societal perspectives. this section embarks on a journey through the historical progression of ADHD's diagnostic framework, exploring the key milestones and shifts that have taken place in its diagnostic criteria. This evolution not only reflects advancements in the understanding of the disorder but also underscores the intricate relationship between medical science, cultural context, and the evolving needs of individuals affected by ADHD.

2.3.1 The Term of Minimal Brain Damage (MBD) (1917-1947)

The term "Minimal Brain Damage" emerged as the first label associated with ADHD symptoms, stemming from the belief that brain damage was the underlying cause of these behaviours. This conceptualisation took shape during a period marked by the global influenza/encephalitis epidemic of 1917-1918, which claimed the lives of over 60 million individuals. Among the survivors, children who had endured these brain infections began exhibiting a range of behavioural and cognitive repercussions. These consequences encompassed characteristics such as inattention, impulsivity, hyperactivity, emotional volatility, and antisocial conduct (Ebaugh, 1923; Hhman, 1922; Stryker, 1925), features that closely align with today's definition

of ADHD. These behavioural patterns not only forged a link between ADHD-like behaviours and brain damage but also prompted researchers to explore possible connections between ADHD behaviours and other forms of brain injury in children. Subsequent investigations revealed that children who had endured brain infections and those exposed to birth trauma, head injuries, and toxins shared similar behavioural issues (Kessler, 1980), thus giving rise to the concept of "brain-injured child syndrome" (Strauss & Lehtinen, 1947).

However, it's important to note that at the time, empirical evidence to firmly establish brain damage as the definitive cause of ADHD behaviours was lacking. This notion was challenged by instances where children without brain injuries displayed analogous behavioural problems. For instance, milder forms of hyperactive behaviours were attributed to psychological factors such as negative child-rearing practices and dysfunctional family dynamics, rather than brain damage (Maniadaki & Kakouros, 2017). This led to the evolution of the term "minimal brain damage" (MBD) into a broader concept that encompassed a spectrum of childhood behavioural and emotional challenges. This shift underscored that while brain injury or damage might correlate with these behaviours, it could not be solely regarded as the exclusive trigger. The 1960s marked a turning point when research findings began questioning the assumption that hyperactivity was universally tied to brain damage. Notably, children with brain damage didn't consistently manifest hyperactive symptoms, and the neurological damage among hyperactive children showed significant variation (Birch, 1964; Rapin, 1964; Laufer et al., 1957), prompting researchers to explore other potential causes of behavioural issues.

In 1957, Laufer et al.'s research shifted the focus by suggesting that dysfunction of the diencephalon might underpin hyperactive behaviours. This theory proposed that functional disturbances, rather than brain damage or injury, might serve as the basis for these aberrant

behaviours. In light of these findings, the term "Minimal Brain Damage" was supplanted by "Minimal Brain Dysfunction" (Clements, 1966). Subsequently, an official definition of Minimal Brain Dysfunction emerged, highlighting three primary symptoms: inattention, impulsivity, and hyperactivity. However, this term proved overly inclusive and heterogeneous, as later research revealed that numerous cases of brain dysfunction didn't exhibit hyperactive behaviours. Moreover, these symptoms correlated with other behavioural disorders such as "learning difficulties" and "language disorders" (Barkley, 2006a; Rothenberger & Neumarker, 2005). This ambiguity, coupled with the term's lack of utility in treatment, led to its abandonment in the 1970s (Barkley, 2015).

The evolving relationship between brain damage and ADHD behaviours eventually introduced medication as a treatment avenue. In 1937, Charles Bradley (1937) reported the positive impact of stimulant medication on various behavioural disorders and learning challenges. While initially intended for treating headaches, Benzedrine (racemic amphetamine) remarkably ameliorated behaviours and enhanced school performance (Laufer, Dehoff & Solomons, 1957). Consequently, stimulants became the primary medication for ADHD treatment across many countries. Nevertheless, recent research has hinted at the limitations of stimulants in improving academic performance for children with ADHD (Pelham et al., 2022).

In summary, the evolution from "Minimal Brain Damage" to present-day ADHD understanding underscores the changing landscape of diagnostic labels and insights. Initially linked to brain damage post the 1917-1918 epidemic, perceptions shifted due to insufficient evidence. Broader perspectives emerged, acknowledging diverse behavioural factors beyond brain damage. "Minimal Brain Dysfunction" and symptom-focused definitions followed, but imprecision led to term abandonment. Concurrently, stimulant medication emerged as a pivotal

treatment, yet recent studies suggest its limited efficacy. This journey reflects the interplay between research, clinical practice, and diagnostic terms, enhancing our comprehension of ADHD's complex nature and treatment approaches.

2.3.2 The Term of Hyperkinetic Impulse Disorder DSM-II (1957-1968)

The way this behavioural disorder was named has been a point of debate due to the uncertainty surrounding its root cause. Researchers have taken a descriptive approach to defining the disorder, moving away from attributing it to a single cause. In the 1950s, the focus of research shifted towards understanding how this disorder develops. Many studies explored changes in the brain's functioning. Some researchers believed that the hyperactive symptoms might result from an excess of brain activity due to difficulties in filtering incoming information in the brain (Laufer, Denhoff, & Solomons, 1957). They gave this version of the disorder the name "Hyperkinetic Impulse Disorder."

At the same time, another line of thinking emerged, suggesting that focusing on specific behaviours like hyperactivity and impulse control might be more useful. Hyperactivity became a key aspect in defining this disorder. This shift marked a move from looking for a single cause to describing the disorder based on its characteristics. The definitions became more detailed. For instance, Chess (1960) defined "hyperactive child syndrome" as "a child who moves faster or more often than the average child, or is always in motion, or both" (p. 2379). This definition emphasised the physical activity associated with the disorder, differentiating it from conditions linked to brain damage. This perspective paved the way for the formal diagnostic criteria of ADHD. Although ADHD-like symptoms were recognised as a distinct condition, a formal diagnosis wasn't established until later. In 1968, DSM-II introduced the term "hyperkinetic reaction of childhood" to describe this disorder. *Table 3* shows the diagnostic criteria for Hyperkinetic Reaction in Childhood in DSM-II.

Table 3 Diagnostic criteria for Hyperkinetic Reaction of Childhood (or Adolescence) in DSM-II

This disorder is characterized by overactivity, restlessness, distractibility, and short attention span, especially in young children; the behaviour usually diminishes in adolescence.
If this behaviour is caused by organic brain damage, it should be diagnosed under the appropriate non-psychotic organic brain syndrome.

Note. (APA, 1968, p50)

The second edition of DSM marked a significant step in establishing standardised language to describe behaviours resembling ADHD. It introduced a method of defining the disorder by detailing its symptoms. However, these descriptions lacked clarity and depth, making them insufficient for accurate diagnosis. As evident in Table 3, the sole identifier of the disorder was "overactivity, restlessness, distractibility, and short attention span." The vagueness and lack of specificity rendered them of limited practical use in clinical settings.

Additionally, these criteria failed to encompass all the key features of the disorder. Subsequent research revealed that two other central traits of ADHD, impulsivity and inattention, were inadequately defined and incorporated into these diagnostic criteria. Some researchers regarded impulsivity as a fundamental characteristic of the behavioural disorder (Wender, 1971; Douglas, 1972). Concurrently, there existed differing opinions on which attributes predominantly contributed to the challenges faced by Hyperkinetic children. Some believed that difficulties in maintaining attention and managing impulses were more pivotal in explaining the struggles associated with this disorder (Douglas, 1972; Werry & Sprague, 1970).

In summary, the evolution of terminology surrounding this disorder reflects a transition from seeking a single cause to embracing a descriptive understanding of its characteristics. This

transition has paved the way for refined diagnostic criteria that better encapsulate the multifaceted nature of the disorder.

2.3.3 DSM-III ADD Attention Deficit Disorder (1980)

Following the release of DSM-II, the depiction of Attention Deficit Disorder (ADD) encountered disputes due to its ambiguous and limited diagnostic utility. This prompted investigations into alternative facets of the disorder. The research focus shifted from hyperactive to inattentive symptoms. Dykman et al. (1971) reimagined "Minimal Brain Damage" as an attentional deficit syndrome, associating short attention spans and delayed reactions with specific learning challenges. Consequently, inattention took centre stage in defining the disorder.

In 1979, Douglas et al. introduced a model of attention deficits encompassing four primary deficits that outlined ADHD symptoms: (1) attention investment, organization, and maintenance; (2) inhibition of impulsive responses; (3) arousal modulation based on situational demands; and (4) an inclination for immediate reinforcement (Barkley & Murphy, 1998). This model exerted significant influence, shaping extensive research and guiding DSM-III (Barkley & Murphy, 1998).

The third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III; APA, 1980, p43-44) marked substantial changes to the criteria. It relabelled symptoms as Attention Deficit Disorder with Hyperactivity (ADDH) and implemented revisions. Noteworthy changes included:

Name change: The disorder became Attention Deficit Disorder (with or without Hyperactivity), reflecting a reconceptualization; Introduction of a descriptive behavioural list: Inattention, impulsivity, and hyperactivity were categorized, with impulsivity gaining prominence for the

first time; Introduction of cut-off scores, symptom duration, onset age, and exclusion criteria: Behavioural checklists for primary symptoms were established, along with specifics on symptom number, duration, age of onset, and exclusion criteria; Introduction of subtypes: ADD-H (without hyperactivity) and ADD+H (with hyperactivity) subtypes were introduced, offering distinctions based on behavioural patterns; Inclusion of criteria for adolescents and adults: Diagnostic criteria for Attention Deficit Disorder, Residual Type (ADD-RT), acknowledged persistent symptoms' potential impact on adolescents and adults' social, academic, and occupational lives.

Despite DSM-III's pivotal contribution and a framework for future diagnostic criteria, the ADHD concept still poses limitations. Attention-centric criteria and controversies persist around the definition and differentiation of subtypes. The interplay and distinctiveness of hyperactivity, impulsivity, and inattention remain subjects of debate (Achenbach & Edelbrock, 1983; Milich & Kramer, 1984).

2.3.4 DSM-III-R ADHD Attention Deficit and Hyperactivity Disorder (1987)

Amid ongoing debates about its dimensions, the revised category adopted a unidimensional structure, featuring a single list (See Appendix 2). The changes in the new criteria include:

•The Disorder's Name Change: Further research emphasised hyperactivity and impulsiveness as defining characteristics of the disorder (Barkley, 1990; Weiss & Hechtman, 1993), prompting its renaming to "Attention-deficit/hyperactivity disorder (ADHD)."

• Unified Behavioural List: Replacing the previous three-category framework, a single list incorporating 14 problematic behaviours was introduced, along with a singular cutoff score.

• Severity Classification: A Mild, Moderate, and Severe rating scale was introduced to gauge the severity of the disorder.

• Shift in Subtypes: The ADD-H subtype transitioned to an undifferentiated attention deficit disorder (UADD) residual category, accompanied by the removal of exclusion criteria for other disorders. The DSM-III-R removed the ADD+H/-H subtypes due to their controversial nature (Barkley, 2006).

In essence, DSM-III-R adopted a developmental framework for ADHD-like behaviour diagnosis, while also delineating boundaries to distinguish the disorder from other mental developmental issues. However, the criteria for UADD and the exclusion of mental development problems remained enigmatic, curtailing their practical application. Furthermore, the unidimensional structure of ADHD criteria proved unreliable in clinical contexts. The threshold of 8 symptoms for diagnosis lacked clear guidance on symptom allocation among the disorder's three main characteristics. Consequently, a child diagnosed with ADHD primarily due to inattention could share the same diagnosis label as a child primarily impulsive or hyperactive. This diagnostic ambiguity hindered the selection of appropriate interventions (Biederman et al., 1997). In conclusion, the unidimensional structure of ADHD criteria in the DSM-III-R, coupled with diagnostic ambiguities and limitations, underscored the need for further refinement to enhance the accuracy and specificity of ADHD diagnosis and subsequent interventions.

2.3.5 DSM-IV (1994)

In the late 1980s, the emphasis on inattention as the central aspect of ADHD was questioned as researchers found that attention deficit problems alone couldn't account for all the challenges children faced during tasks. Instead, response inhibition and motor control issues emerged as more reliable explanations for the disorder (Barkley, 2006). Contrary to initial assumptions, studies indicated that hyperactive children didn't necessarily exhibit symptoms of inattention (Van der Meere & Sergeant, 1988). Moreover, the research didn't find deficits in the perception,

filtering, and processing of attention-related information in ADHD children (Maniadaki & Kakouros, 2017). However, significant impacts from motivational factors were observed on ADHD symptoms (Glow & Glow, 1979), highlighting the role of motivation and deficits in reinforcement. This perspective explained major issues faced by ADHD children, such as a lack of self-regulation and consideration of consequences (Benninger, 1989; Quay, 1988).

Meanwhile, the evolution of technology brought about novel neurological and genetic tools (PET, MRI) employed in ADHD research. These studies delved into ADHD from inheritable, genetic, and neurological angles (Zametkin et al., 1990; Biederman et al., 1995), compelling the need for updated definitions and diagnostic criteria. In 1994, the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; APA, 1994) was published, designating the syndrome as Attention-Deficit/Hyperactivity Disorder (AD/HD) (See Appendix 3). Leading experts in the field formed a committee before publication, conducting a literature review, informal survey, and trial to shape these criteria. Based on an informal survey of rating scales assessing behavioural dimensions related to ADHD and a field trial involving 380 children from 10 sites in North America, the criteria were empirically derived (Lahey et al., 1994). These criteria exhibited substantial improvements over their predecessors (Barkley & Murphy, 1998). The changes in the new criteria include:

• A descriptive symptoms list replaced the singular list with two categories. The descriptive symptoms list was divided into two parts: one with 9 items for inattention symptoms and another with 9 items encompassing hyperactivity (6 items) and impulsivity (3 items). These two categories mirror the two dimensions of the disorder: hyperactivity and impulsivity, and inattention. This separation aids clinical diagnosis of disorder subtypes.

• Three subtypes were introduced: predominantly inattentive, predominantly hyperactive-impulsive, and combined.

• The importance of cross-setting diagnosis (home, school, work) was emphasized.

DSM-IV has been widely employed for diagnostic and research purposes, with its criteria demonstrating reliability and validity in identifying individuals with ADHD within populations (Willcutt et al., 2012). Yet, ongoing debates persist surrounding these criteria. First, the question of whether ADHD-I functions as a distinct subtype or shares common attention deficits with other types remain unresolved. Examination of children's classroom behaviours via factor analysis has suggested the presence of two types of inattention: an inattentive-passive form and another related to persistence and distractibility (Achenbach, 1986). Additionally, Lahey et al. (1988) identified a third inattention dimension in clinical ratings known as "sluggish tempo." This implies potential differences between ADHD-I and inattention within ADHD-C. The diverse inattentive behaviours correlate with distinct behavioural issues; ADHD-I primarily struggles with the selective attention component, while ADHD-C involves distractibility and persistence difficulties (Barkley, 2006). As a result, researchers have argued for the separation of ADHD-I and Combined subtypes into two distinct disorders rather than mere subtypes of a singular condition (Diamond, 2005; Milich, Ballentine, & Lynam, 2001).

Secondly, the classification of ADHD-HI as a separate type from ADHD or merely an earlier developmental stage remains unclear. Research has indicated that hyperactive and impulsive behaviours often precede inattention by several years (Egger, Kondo & Angold, 2006). ADHD-HI predominantly involves preschool-age children and its symptoms tend to decrease with age. Conversely, ADHD-C is more prevalent among school-age children, with inattentive symptoms becoming more pronounced as they grow older (Lahey, Pelham, Loney, Lee & Willcutt, 2005). Hart et al. (1995) noted that hyperactive items in DSM-IV criteria lose sensitivity to the disorder with age, whereas inattention symptoms remain suitable for

identifying the disorder across middle childhood and early adolescence. These subtypes may reflect distinct developmental stages of the same ADHD type rather than separate subtypes. Within clinical practice, instances arise where hyperactive-impulsive symptoms wane in adolescents and adults. Consequently, these individuals may no longer fulfil the criteria for the Combined type, instead fitting the P-inattentive (PI) type. Nevertheless, whether they are qualitatively different from individuals who have exclusively exhibited PI symptoms since childhood remains uncertain. Hence, the DSM-IV's definitions of ADHD subtypes continue to pose challenges in clinical applications.

In conclusion, the evolution of ADHD's understanding from its initial conceptualization to the DSM-IV criteria exemplifies the intricate nature of the disorder's diagnostic landscape. The transition from "Minimal Brain Damage" to "Attention-Deficit/Hyperactivity Disorder" reflects the dynamic interplay of scientific advancements, changing perspectives on symptomatology, and the continuous pursuit of diagnostic precision. The journey reveals the complexities of ADHD, its multifaceted symptom dimensions, and the ongoing need for refined diagnostic criteria to guide effective clinical interventions and further unravel the mysteries surrounding this intricate disorder.

2.3.6 DSM-V (2013)

The DSM-V edition, published in 2013 (see Appendix 4), introduces several significant changes to the diagnostic criteria for ADHD:

• The removal of three subtypes and their replacement with three distinct presentations reflects the evolving understanding that ADHD symptoms are dynamic throughout an individual's lifespan, rather than fixed traits. The transition from one presentation to another, such as Combined to Predominantly Inattentive, underscores the fluid nature of these symptoms

(Hurtig et al., 2007). This shift in terminology, from "subtype" to "presentation," acknowledges the variability of symptoms over time.

• The modification of the age onset criterion from 7 to 12 years is based on research indicating that the onset of symptoms before age 12 is more relevant than a specific age. While symptoms need to appear before age 12, impairment does not necessarily hinge on this requirement (Faraone et al., 2006; Guimarães-da-Silva et al., 2012; Todd et al., 2008). Varying accounts of symptom onset from different studies led to this adjustment (Applegate et al., 1997; Kessler et al., 2005), as research suggests that the age of onset does not significantly impact the disorder's characteristics into adulthood (Biederman, Mick, & Faraone, 2006).

• The adjustment of the symptom threshold for older adolescents and adults from 6 to 5 reflects the growing recognition that lower thresholds are more accurate in identifying this demographic (Barkley, 2012; Sandra Kooij et al., 2005).

• The inclusion of three severity specifiers allows for a more nuanced categorisation of the disorder's severity. These specifiers offer mild, moderate, and severe levels of classification.

• The addition of symptom examples aids in clarifying symptom descriptions for older adolescents and adults.

• The requirement for multiple informants, as opposed to a suggestion in DSM-IV, reflects the need for diverse perspectives in diagnosing ADHD and enhances the validity of the diagnosis (Sayal & Goodman, 2009; Sayal & Taylor, 2005; Valo & Tannock, 2010).

• A notable change involves the acceptance of Autism Spectrum Disorder (ASD) as a comorbidity with ADHD. Unlike DSM-IV, ADHD can now be diagnosed alongside ASD, acknowledging the complexity of symptom overlap (Tannock, 2012).

These changes in DSM-V's ADHD diagnosis criteria have the potential to enhance diagnostic reliability. Shifting the focus from impairment evidence to symptom pervasiveness and age

onset reflects the understanding that symptoms can exist without impairment, but not the other way around. This shift may lead to increased diagnostic reliability, but it also raises concerns about potentially broadening the ADHD diagnosis without corresponding impairment (Ghanizadeh, 2013). The introduction of severity levels attempts to refine diagnosis, but its subjective nature and variability across contexts limit its reliability. Overall, while these changes provide a more comprehensive framework for understanding ADHD, their practical implementation and impact on clinical practice remain subjects of ongoing discussion and exploration.

2.3.7 Comparison of ADHD Diagnosis Criteria between DSM and ICD

Both DSM-5 and ICD-11 play crucial roles in facilitating consistent and standardized diagnosis of ADHD across different countries and healthcare systems. However, the selection of which criteria to use often depends on historical, cultural, and practical factors. These factors contribute to variations in how ADHD is conceptualised and diagnosed across different regions.

The historical development of diagnostic criteria played a significant role in shaping the approach different countries take. For instance, the DSM system, developed by the American Psychiatric Association (APA), has undergone several revisions over the years, reflecting changes in diagnostic understanding and clinical practices primarily within the United States. On the other hand, the ICD system, developed by the World Health Organization (WHO), has been used internationally for various medical and health-related purposes. The differences in the origins and objectives of these systems can lead to variations in their adoption for ADHD diagnosis. Cultural perspectives on mental health and behaviour can influence how ADHD is perceived and diagnosed. Different cultures may have varying beliefs about what constitutes "normal" behaviour in children and how behavioural challenges are interpreted. These cultural

factors can impact how ADHD symptoms are recognised and categorised, leading to differences in diagnostic criteria preference.

Other practical factors, such as professional traditions, healthcare system regulations, research and clinical practices and continuity and tradition also influence the use of diagnostic criteria. For example, clinicians in North America might be more inclined to use the DSM criteria, while those in Europe may be more accustomed to using the ICD criteria. Some countries have regulations that guide the use of specific diagnostic criteria for reimbursement and treatment purposes. This can impact which criteria are favoured for diagnosis. If a particular set of criteria is more supported by local research or is aligned with clinical practices, it might be more widely adopted. Once a diagnostic system is established within a country, it can become a matter of tradition and continuity. In summary, the reasons for different countries using different diagnostic criteria for ADHD are multifaceted and involve historical, cultural, professional, and practical factors. These factors contribute to the diversity of perspectives and approaches to diagnosing ADHD across the globe.

DSM-5 and ICD-11 are two prominent classification systems used to diagnose ADHD, and while they share commonalities, there are notable differences between them. First, the name of the disorder differs in the two sets of criteria. In the DSM, it is termed Attention Deficit and Hyperactivity Disorder (ADHD), while the ICD refers to it as Hyperkinetic Disorder (HKD). While the primary symptoms may vary, with DSM focusing on inattention, hyperactivity, and impulsivity, and ICD emphasising impaired attention and overactivity, the lists of symptoms themselves are notably similar.

Second, a significant distinction between the two criteria lies in their approaches to establishing a diagnosis. When compared to DSM-V, ICD-10 has more stringent rules and a higher threshold for diagnosis. The ICD-10 criteria mandate the observation of a specific number of symptoms along with evidence of related impairments in each setting to establish a diagnosis. Conversely, DSM-V criteria concentrate solely on the number and duration of symptoms. This difference contributes to potentially lower prevalence estimates in ICD-10. Research by Polanczyk et al. (2014) demonstrates that using ICD-10 leads to average prevalence estimates 6.4% lower than those derived from DSM-IV.

Third, ICD offers two versions tailored to different users: the "blue book" for general clinical, educational, and service use, and the "green book" designed for research purposes that offers an extended framework for scientific investigation (World Health Organization, 2010). Despite the adoption of distinct criteria in different continents, there is a shared agreement on the symptoms. DSM criteria have gained wider global adoption and hold greater influence than ICD criteria (World Health Organization, 2010). To establish standardised criteria and diagnosis procedures, the updated ICD-11 criteria have been aligned with DSM-V, resulting in the replacement of Hyperkinetic Disorder with ADHD.

2.4 Conclusion

As discussed earlier, despite the formal diagnosis of ADHD being established around a century ago, its behavioural symptoms and related impairments have captured attention for centuries. With advancements in modern medicine, efforts have been directed towards identifying the cause of this childhood behavioural issue and formulating an accurate diagnosis. Initially deemed a moral deficiency in children, discipline was the recommended remedy. Subsequently, biological antecedents beyond individual control, such as brain damage and dysfunction, were thought to underlie this disorder. Terms like Minimal Brain Damage, Brain Dysfunction,

Hyperkinesis, and Hyperkinetic Reaction were used to describe it. However, further research revealed that children without such biological factors also displayed symptoms, shifting the focus towards a descriptive understanding of the core behavioural symptoms. This shift also signifies that this disorder is characterised by observable behaviours rather than blood tests or brainwave analysis, resulting in a dimensional rather than strictly categorical diagnosis.

In light of this shift, significant attention has been dedicated to defining, assessing, and tracing the life course of ADHD. Professionals in the field have diligently worked to establish diagnostic criteria, notably through the DSM and ICD systems. Three primary characteristics—inattention, impulsivity, and hyperactivity—have been identified as markers of this behavioural condition. As explored in preceding sections, criteria have evolved from DSM-III, which laid out a list of 16 behavioural symptoms delineating the disorder. Subsequent iterations of the criteria primarily focused on shaping the dimensions of this behavioural issue, elucidating the relationships between these dimensions, associated impairments, the earliest age of diagnosis, manifestation duration, and diagnostic thresholds. Throughout these changes in ADHD diagnosis criteria, the behavioural list has remained at the core, providing a standardised method for professionals across diverse geographical regions to identify and diagnose the condition. This categorical definition facilitates global ADHD diagnosis and enables the comparison of prevalence rates across different countries.

Yet, as these diagnosis criteria are implemented widely, a trend of increasing prevalence worldwide, alongside a wide range of prevalence rates, has emerged. This suggests that other factors are influencing diagnosis outcomes. Diagnosis is founded on the observation of 18 behavioural symptoms, considering factors like the number and duration of manifested symptoms, symptom severity, and related impairments. This behavioural perspective implies

that specific symptom definitions must be contextualised (Goldstein, Goldstein, Jones, Braswell & Sheridan, 1998). Consequently, numerous contexts come into play when observing these behavioural symptoms, encompassing environment, culture, and social influences.

Where are these behavioural symptoms observed—school, home, or clinical settings? Children's conduct varies across different environments. Who observes and assesses these behaviours—parents, teachers, or peers? Different observers possess varying perceptions and beliefs about children's conduct. Hence, the 18 behavioural symptoms are subject to multiple dimensions of observation (environment, culture, and social factors). These dimensions within the diagnostic criteria might contribute to the variation in ADHD prevalence across countries and even within a single country.

Timimi (2005) proposed that the escalating prevalence of ADHD could arise from shifts in how society perceives children's behaviour and changes in the environment. Behaviours once seen as typical or interpreted differently are now regarded as indicative of a behavioural issue. Such shifts in perception could potentially alter prevalence rates. Furthermore, environmental transformations may influence the development of ADHD among children (Livingstone et al., 2016). Thus, delving into and comprehending the environmental, social, and cultural aspects of ADHD diagnosis play a pivotal role in enhancing the understanding of the disorder. Cross-cultural research presents a valuable avenue to explore how different cultures interpret these diagnostic criteria, thereby identifying potential cultural influences underlying ADHD diagnosis outcomes. The next chapter will delve into the details about the role of culture and raters play in the diagnosis of ADHD.

Chapter 3 Understanding ADHD Diagnosis and the Role of Culture and Raters

3.1 Introduction

Understanding the historical context and diagnostic advancements is crucial for gaining insights into the complexities of ADHD diagnosis. Furthermore, as I mentioned in the previous chapter, cross-cultural research offers us a spectacular goal to explore how these diagnostic criteria are perceived by different cultures. The study aimed to propose a novel approach to generate a relatively fair comparison of children's ADHD prevalence between culturally distinct groups, emphasising the importance of considering cultural and rater factors in ADHD diagnosis. Therefore, it is essential to explore how the diagnosis is carried out and related issues in the diagnosis procedure in different cultural contexts. By examining the diagnostic practices and perspectives in different cultural contexts, I can better understand the variations and similarities in ADHD diagnosis.

This chapter provides a comprehensive exploration of various topics related to ADHD diagnosis, including the general procedure, the role of raters, the impact of culture, cross-cultural comparisons, and specific contexts such as the UK and China. By addressing these topics, this chapter contributes to the broader goal of enhancing the accuracy and fairness of ADHD diagnosis across diverse populations.

3.2 The General Procedure of ADHD Diagnosis

The diagnosis of ADHD in children is a complex process, as numerous biological and psychological issues can give rise to symptoms resembling ADHD. These issues may either be the actual cause of ADHD-like behaviours or coincidentally coexist with ADHD. To achieve an accurate diagnosis, these potential underlying causes must be taken into account. As a result, there is no single definitive test that can definitively determine whether a child has ADHD or not. The diagnostic approach relies on behavioural assessments, employing multiple methods

of data collection across various settings and informants (Barkley, 2006). The diagnostic process is comprehensive and multidimensional, lacking a universally standardised set of guidelines for its administration, given variations in medical systems across countries.

Taking the UK as an illustrative example, the diagnostic process encompasses two key stages: initial screening evaluation and formal assessment (NHS, 2018). In the initial screening evaluation, a child showing signs of ADHD receives an assessment from a local GP. This evaluation involves gathering information about the symptoms, their onset, occurrence settings, impact on the child's life, family history of ADHD, the child's overall health, and significant life events. Based on this information, the GP forms a preliminary conclusion about whether ADHD may be a possibility.

However, it is important to note that formal diagnosis cannot be established by the local GP alone. A child who is deemed to potentially have ADHD through the initial screening is referred for a formal assessment. This comprehensive assessment encompasses biological, psychological, and social evaluations to rule out alternative causes of the symptoms. Specialists such as child psychiatrists, paediatricians, learning disability experts, social workers, and occupational therapists with ADHD expertise participate in this assessment. They conduct psychiatric examinations to eliminate other potential causes and engage in interviews with the child, parents, and teachers to gather in-depth information regarding symptoms and related impairments in various contexts.

This diagnostic process highlights that ADHD diagnosis is not a single test-based determination but rather a procedure involving information gathering and interpretation. Despite guidelines like DSM-V and ICD-11, the symptom lists do not encompass all

dimensions of the disorder. The symptoms of ADHD are multidimensional, and their presentation varies among different children, introducing some degree of ambiguity into the diagnostic process. In general, children typically receive an ADHD diagnosis due to concerns expressed by parents or teachers regarding potential behavioural problems that could hinder the child's functioning at home, school, or other settings.

Although medical specialists in ADHD have undergone comprehensive training, and clinical practice, and gained experience, they do not directly observe the child's behaviours at home or school. Their assessments heavily rely on information from parents and teachers. Teachers and other school staff often play a crucial role in initiating ADHD diagnoses in children (Sax & Kautz, 2003). Furthermore, teacher ratings of children's ADHD behaviour are the most frequently employed methods for school-based ADHD screening (Barkley, 2006). Therefore, understanding the vital importance of teachers in diagnosing ADHD in children is essential.

3.2.1 Criticism of the Diagnosis of ADHD

The diagnosis of ADHD in children has drawn criticism due to the ambiguity inherent in its definition and its susceptibility to varied interpretations. ADHD is characterised by categories of symptoms, stemming from sources like statistics, observation, education, behavioural research, and experiments. These symptoms, however, are not exclusive to children with 'ADHD'; they represent broader categories of child behavioural difficulties rather than distinct markers of ADHD itself (Baughman, 2006; Mather, 2012). Critics emphasise that when taken individually, these ADHD symptoms often mirror instances of typical childhood behaviour that ordinarily wouldn't lead to distress or impairment (Baughman, 2006; Mather, 2012).

These critics argue that the diagnosis of ADHD relies on the combination and frequency of otherwise "typical" observations. Children diagnosed with ADHD are those whose behavioural

challenges raise concerns among parents or teachers. Unlike physical ailments, there are no specific physical symptoms, cognitive, metabolic, or neurological markers, blood test results, brain scan findings, or chemical imbalances that can definitively confirm ADHD in a child (Breggin, 1998). This makes differentiating these symptoms from ordinary childhood behaviours a challenging endeavour (Williams, 2008).

Moreover, the diagnosis heavily depends on subjective judgments from various parties, including parents, doctors, psychologists, and teachers, to ascertain the presence of ADHD. Research has shown that agreement among these raters is relatively low (O'Neill et al., 2013), casting doubt on the reliability of a diagnosis based on such subjective assessments. The overlap of symptoms with other conduct disorders further complicates the precise identification of ADHD (Timimi, 2002). The growing global prevalence of ADHD has raised concerns about the overmedication of individuals to conform, rather than acknowledging their unique differences and addressing their individual needs. A false negative diagnosis could result in an individual with ADHD missing out on essential support and treatment for their associated challenges. Conversely, a false positive diagnosis comes with risks such as unnecessary long-term medication, side effects, financial burdens for families and society, and the stigmatisation of the child. As such, understanding the factors influencing ADHD diagnosis and striving for improved accuracy becomes pivotal.

Given that the diagnosis hinges on subjective assessments from various informants across diverse contexts, their beliefs about ADHD and cultural backgrounds inevitably taint their judgments on what constitutes symptoms. Varying perspectives on ADHD shape how individuals perceive the diagnostic process. Despite the global reporting of ADHD symptoms and extensive research, substantial debates persist on whether ADHD is indeed a disorder and

how it should be diagnosed and treated. Amid these debates, the identification, referral, diagnosis, and treatment of children with ADHD are significantly influenced by the knowledge and perspectives of all parties involved (Hamed et al., 2015).

Batchelder (2003) asserted that social and cultural factors play a crucial role in understanding the disparities in ADHD diagnosis and medical interventions. These factors can shape the perception of ADHD-like behaviours as problematic. Given that inattention and impulsivity are inherent human behaviours, their perceived level of impairment is highly contextual. Children, being in a state of development, learn to regulate their attention and impulsivity within social contexts, acquiring an understanding of acceptable behaviour. Children do not independently seek help or perceive themselves as having issues; instead, they rely on external individuals, like parents and teachers, to identify their condition. Despite efforts towards standardisation, cross-cultural studies reveal significant variations in how symptoms of ADHD are rated among raters from different countries. Consequently, grasping the impact of raters and cultural factors is fundamental to the ADHD diagnosis process.

3.3 The Raters' Role in the Diagnosis of ADHD

The raters play a crucial role in the diagnosis of ADHD, as they are the ones responsible for observing and evaluating the behaviours exhibited by children. This section focuses on understanding the significance of raters, particularly parents and teachers, in the ADHD diagnosis process.

3.3.1 Parents' Role in the Diagnosis of ADHD

As emphasised earlier, behavioural rating is a widely used approach in diagnosing a child with ADHD. Typically completed by individuals who have regular interactions with the child, such as parents and teachers, this approach serves as a primary diagnostic tool. Parents, in particular, offer valuable insight into identifying a child's ADHD symptoms, especially within the home

environment. In certain scenarios, parents are the sole source of information, particularly when teachers are new or unavailable, or when parents choose not to involve or inform the school. Nevertheless, some researchers have raised concerns about potential bias in parent ratings.

To begin with, the validity of parental ratings might be influenced by their own symptomatology. Familial aggregation studies have supported the notion of a genetic contribution to ADHD, suggesting that children and their parents might exhibit shared symptoms (Epstein, 2000; Faraone et al., 1994). Consequently, when parents with ADHD themselves rate their children's symptoms, their assessments could be skewed due to the likelihood of them experiencing similar symptoms.

Secondly, parent ratings of their child's ADHD symptoms within a school setting have been called into question due to varying observations. Discrepancies often arise between parent and teacher ratings regarding ADHD symptoms in the school context. Mitsis et al. (2000) discovered that parents tended to underestimate the extent of hyperactive symptoms exhibited by their children at school, leading to an under-identification of ADHD cases within the school environment. Conversely, another study suggested that parents reported more hyperactive and impulsive behaviours compared to teachers for younger children (Papageorgioun et al., 2008). Further research indicated that while parents of children reported exhibiting high symptom levels at school tended to rate such symptoms at home, only half of these children met the diagnostic criteria (Sayal & Goodman, 2009). Sayal & Taylor (2005) found that parental assessment of their children's hyperactive symptoms at school was influenced by their own mental health and the child's behaviour within the home.

The disparity between parent and teacher ratings might be attributed to parents' limited experience in benchmarking their child against a larger peer group, leading them to potentially exaggerate minor issues. Furthermore, parental perception of their child's problematic behaviours can be shaped by their parenting style, impacting their ratings and the manner in which they engage with rating scales. Parents may view the rating scale as a collection of common everyday problems rather than a clinical compilation of ADHD symptoms. As a result, parents may magnify minor issues within the home setting while underestimating actual challenges within the school environment. Nonetheless, parents' evaluation of their child's ADHD symptoms remains a valuable information source; nevertheless, it is crucial that an ADHD diagnosis is grounded in data gathered from multiple sources across diverse settings.

3.3.2 Teacher's Role in the Diagnosis of ADHD

Teacher ratings constitute another invaluable source of information. Derks et al. (2007) discovered that teachers reported higher levels of disruptive behaviours at school for boys compared to girls, in contrast to mothers' ratings which did not display gender differences. Additionally, findings from an ODD study indicated that teachers reported higher prevalence rates for boys than girls, a disparity not observed in parents' reports (Meisel et al., 2013). This discrepancy might stem from the more demanding nature of the school environment, causing gender-related distinctions in ADHD behaviours to be accentuated in school settings. Hence, teacher ratings offer essential insights into children's problematic behaviours within different contexts.

Furthermore, teachers are often regarded as ideal informants for reporting children's challenging behaviours in comparison to parents. First, in contrast to parents, teachers observe children in structured and demanding settings for extended periods, facilitating a comprehensive assessment of behaviour across diverse activities and a broad cohort. Second,

teachers possess professional training and knowledge of child development, enhancing their awareness of problematic behaviours. Survey results indicate that approximately half of all teachers have experience teaching children with ADHD, with most educators feeling confident in their ability to teach such children, whether or not they have a clinical diagnosis (Sayal et al., 2006). Consequently, armed with professional training and experience, teachers are well-equipped to differentiate children struggling with ADHD symptoms from their typically developing peers.

Additionally, the increased stress and negative interactions that teachers often experience with children displaying ADHD behaviours motivate them to recognise these symptoms and seek external support (Barkley, 2006; Greene et al., 2002). The research highlights parents' belief that teachers provide better advice on childhood psychiatric issues than psychiatrists, attributing this perception to teachers' daily interactions with children. Furthermore, teachers are frequently the first individuals consulted by parents when they notice behavioural problems in their children (Pescosolido et al., 2008; Sayal et al., 2006). As a result, teachers play a pivotal role in identifying and referring children with ADHD, rendering them essential contributors to this process.

3.3.3 Factors Affect Teachers' Rating of ADHD

However, several factors can influence teachers' assessment of ADHD symptoms. Teacher judgment is rooted in their observations of a child's behaviours within the classroom context. Jacobson (2002) asserts that teachers' evaluations of hyperactive, impulsive, and distractible behaviours are confined to specific environments, suggesting that their assessments of appropriate and inappropriate behaviours are highly context-dependent. This contextual aspect stands in contrast to the diagnostic criteria (DSM-V, ICD 10), which offer only general descriptions devoid of contextual details. This discrepancy provides room for teachers to

interpret how frequently certain behaviours align with the ADHD symptom criteria within specific contexts, contributing to the potential variability in teacher ratings. Consequently, multiple factors can impact teacher assessments, including their familiarity with ADHD, teaching experience, student gender, grade level, classroom dynamics, school quality, and geographic location (urban/rural).

To begin, teachers' knowledge of ADHD profoundly influences their judgments. Research indicates that many teachers acquire their understanding of ADHD through interactions with students exhibiting these behaviours in their classes (Kos et al., 2004; Sciutto et al., 2000). However, not all teachers have experience with students diagnosed with ADHD. Furthermore, exposure to a child with ADHD may only confer familiarity with certain symptoms, potentially leaving other disorders like ODD and ASD, which share symptoms with ADHD, unrecognised. Notably, there is a significant lack of knowledge about and misconceptions about ADHD among teachers worldwide (Perold et al., 2010; Kos et al., 2004; Sciutto et al., 2000; Snider, Busch & Arrowood, 2003), influencing their interpretation of ADHD symptoms in the classroom setting. While Stevens et al. (1998) argue that factual knowledge of ADHD might not affect the accuracy of teacher ratings, as these ratings primarily concern identifying symptoms rather than understanding the disorder's causes and treatments, the ability to recognise and interpret behaviours as related symptoms necessitates a comprehensive understanding of ADHD symptoms within their specific contexts. Consequently, this knowledge does indeed impact the accuracy of teacher ratings.

Furthermore, a teacher's teaching experience not only shapes their perception of student behaviour in the classroom but also influences their areas of focus. Borg (1998) found that more experienced teachers tended to view problematic behaviours with less concern than their

less experienced counterparts. Additionally, Kokkinos et al. (2002) discovered that less experienced teachers exhibited greater concern for students' externalising behavioural issues, while their more experienced colleagues were more attentive to internalising behavioural problems. However, another study revealed that professional experience with children diagnosed with ADHD could lead to a more tolerant view of students' behavioural challenges, potentially introducing bias into teachers' judgments (Stevens & Quittner, 1998).

Moreover, teachers' perceptions might be affected by the gender of the students. Teacher ratings reflect a gender disparity in the presentation of symptoms, with boys often reported to exhibit more hyperactive and impulsive behaviours than girls in school (Arbuckle & Little, 2004; Wheldall & Merrett, 1988). Nevertheless, it has been argued that this gender discrepancy arises from differing perceptions among teachers (Jackson & King, 2004). Additionally, students' grade levels can impact teachers' assessments. As students advance in grade levels, fewer cases of disruptive behaviours like talking out of turn might be reported, while inattention behaviours could become more prevalent (Houghton et al., 1988). Furthermore, even for the same symptom, teachers from different grade levels might perceive it differently. Ding et al. (2008) found that high school teachers considered talking out of turn less problematic compared to their counterparts in middle and elementary schools.

Beyond this, classroom settings also shape teachers' perceptions of student behaviour. Factors such as seating arrangements, subject matter, and individual versus group work have been identified as influences on teachers' interpretations of student behaviours (Turnuklu & Galton, 2001). According to research by Houghton et al. (1988), modern language teachers reported higher instances of students exhibiting problem behaviours, while craft design technology and remedial course teachers reported fewer such instances. Furthermore, Leblanc et al. (2007)

found that though the bulk of the variance in teacher reports of behaviour problems could be attributed to differences between teachers within a given school, teachers from urban high schools reported significantly more behaviour problems compared to those from rural schools.

Moreover, significant discrepancies exist in teachers' perceptions and interpretations of students' classroom behaviours across different cultures. Brewis et al. (2000) identified notable differences between American and Colombian teachers in how they construed and managed students' behaviours in the classroom. Additionally, Jacobson (2002) found distinctions between English and American teachers in their criteria for appropriate behaviours. American teachers tended to rate children higher on hyperactive and impulsive behaviours compared to their English counterparts. This variance might stem from differences in teachers' standards for appropriate behaviours, influencing their teaching methods and, in turn, whether children appear distracted or hyperactive. Hence, comprehending these variations in teachers' perceptions is crucial for accurately assessing ADHD symptoms.

In summary, teachers' assessment of ADHD symptoms is a multifaceted process shaped by numerous factors. Grounded in classroom observations, this judgment is contextual and contrasts with the broad diagnostic criteria. Teachers' ADHD knowledge, gained through interactions with students, varies and can leave them unfamiliar with certain disorders sharing symptoms with ADHD. A global knowledge gap and misconceptions about ADHD also affect symptom interpretation. Teaching experience influences concern levels and focuses on behaviours, while student gender, grade level, and classroom dynamics contribute to differential perceptions. Cultural variations further compound these differences, highlighting the necessity of understanding these complexities for accurate ADHD symptom assessment in diverse educational contexts.

3.3.4 Inter-rater Agreement among Informants in the Rating of ADHD

In the realm of Attention-Deficit/Hyperactivity Disorder (ADHD) diagnosis, the assessment of symptoms through different informants, such as parents and teachers, plays a crucial role in forming a comprehensive understanding of a child's behavioural profile. However, the extent to which these informants agree on the presence and severity of ADHD symptoms is a crucial aspect of diagnostic accuracy and reliability. The phenomenon of Inter-rater Agreement among informants holds significance in assessing the consistency of ADHD symptom perception across diverse contexts.

Inter-rater Agreement stands as a crucial measure for assessing the consistency of raters when utilising a common scale to evaluate identical targets (Shweta et al., 2015). It is an integral facet of the reliability and validity of any rating scale, offering insights into the extent of consensus and homogeneity among various raters' judgments. Essentially, Inter-rater Agreement demonstrates the reliability of raters' utilisation of the rating scale while also indicating its validity in accurately measuring the intended trait. In cases where raters exhibit discordant perspectives, it raises questions about the scale's design's appropriateness, potentially causing raters to interpret and apply it differently. Alternatively, it could signify a lack of adequate training for raters in using the scale. Both scenarios pertain to test validity concerns, which hold significance in the assessment of any rating scale.

While both DSM-V and ICD-11 require the involvement of multiple informants from diverse contexts for ADHD diagnoses, the degree of inter-rater agreement among these informants remains modest. ONeill et al. (2013) observed minimal agreement not only between teachers and parents (Kappa = -.16) but also between teachers and clinicians (Kappa = .13) regarding ADHD behaviours in preschool children. Similarly, Murray et al. (2007) reported correlations

between parents' and teachers' ratings of ADHD symptoms ranging from .24 to .26. Moreover, the level of agreement seems to be contextually influenced. Sollie et al. (2013) noted that the agreement between mothers and fathers on children's ADHD behaviours is higher than the agreement between parents and teachers. Thorell et al. (2018) found that informants within the same setting (e.g., both parents or pairs of teachers) exhibit an agreement level of .60, whereas informants from different settings (e.g., parents and teachers) demonstrate an agreement of .28.

These findings indicate that the agreement among informants from similar settings tends to be higher compared to those from diverse settings. This phenomenon might arise from the context-specific nature of ADHD symptoms (Hartman et al., 2007; Gomez, 2007). Accordingly, parents, teachers, and clinicians could observe distinct ADHD behaviours across varying contexts, possibly due to the child's differing behaviours in different settings. Consequently, this context-driven variability contributes to the observed low to moderate inter-rater agreement among multi-informants across varying settings for ADHD symptom ratings. However, this observation does not fully explain the relatively modest agreement even between informants within the same setting, such as mothers and fathers. Duhig et al. (2000) discovered that the agreement between mothers and fathers regarding ADHD symptoms ranged from .38 to .40, based on an intraclass correlation meta-analysis encompassing 60 studies.

Despite the diagnostic criteria requiring multi-informant involvement across diverse settings, they do not demand the same setting's multi-informant collaboration. While numerous studies have examined the agreement between parents' ratings of children's ADHD symptoms, few have delved into the consensus among teachers' ratings of such symptoms. Given that individual teachers may possess distinct interpretations and tolerance thresholds for these behaviours, concerns have emerged regarding the reliability of relying on a solitary teacher's

judgment to assess children's ADHD behaviours within the school environment. Therefore, it is essential to investigate the IRA among teachers when they rate children's ADHD behaviours.

3.4 The Role of Culture in ADHD Diagnosis

Children with ADHD are recognised in every cultural context (Barkley et al., 2002). However, as discussed in the previous chapter of History of Aspects of ADHD, it becomes evident that the widely utilised diagnostic criteria for ADHD, namely the DSM or ICD, were formulated by Western research, employing Western population samples, within the framework of Western culture. While these criteria have been applied in diverse cultural contexts, cross-cultural research frequently uncovers variations. Mann et al. (1992) examined disparities in clinical professionals' assessments of ADHD behaviours across four cultural contexts (China, Japan, Indonesia, and America). Their findings indicated that Chinese and Indonesian professionals assigned higher ratings to hyperactive and disruptive behaviours compared to their Japanese and American counterparts. This illustrates that, even when employing uniform criteria and assessing the same targets, perceptions of hyperactive behaviours diverge significantly across distinct cultural contexts. "*The characterisation of behaviours or abilities is a cultural process: basic concepts used to classify people such as normal, disordered, abnormal, and average are culturally constructed and culturally variable*" (Jacobson, 2002, p. 284). To grasp the differences in ADHD diagnosis among various cultures, it becomes essential to explore the role of culture in the process of ADHD diagnosis.

3.4.1 Culture in Teachers' Perspective of Judging Children's Behaviours

The influence of culture on teachers' perspectives when assessing children's behaviours is a significant aspect to consider. Culture shapes how individuals perceive and interpret behaviours, thereby affecting the judgments teachers make regarding students' actions. Studies suggest that teachers from diverse cultural backgrounds may interpret children's ADHD behaviours in varying ways. Brewis et al. (2000) uncovered notable distinctions in how

children's behaviours are understood and managed in classrooms between American and Latin teachers. Specifically, Latin teachers exhibited more positive responses to active and noisy children compared to their American counterparts. This variance is attributed to differences in classroom dynamics, with Latin classroom activities being less linear and valuing increased movement, interaction, and vocalization. Consequently, behaviours akin to ADHD characteristics, which might be viewed as troublesome within American classrooms, could adopt a neutral or potentially positive interpretation within the Latin context, where such energetic engagement is valued.

Further exploration of cultural perspectives on ADHD was conducted by Moon (2011), who compared teachers' viewpoints on ADHD between Korea and America. The research demonstrated that American teachers did not associate students' disruptive behaviours with a reflection of the teacher's competence, unlike Korean teachers, who emphasised their role in disciplining students. Korean teachers often interpreted ADHD students' disruptive actions as a challenge to their authority, potentially regarding these behaviours as transgressions against their control. This tendency could lead to interpreting ADHD symptoms more as behavioural issues rather than clinical disorders.

Another cross-country investigation evaluating teachers' assessments of ADHD behaviours in young children across England, Scotland, and Australia revealed that Scottish teachers perceived their students' inattentive, hyperactive, and impulsive behaviours to be more severe than their counterparts in England and Australia (Merrell et al., 2013). These cross-cultural findings highlight the reality that identical behaviours are subject to differing interpretations within distinct classroom cultural contexts or even present diversely among children from

different cultural backgrounds. As a result, further research is needed to delve into these factors and their implications for the diagnosis of ADHD.

3.4.2 Cross-cultural Studies of Rater Bias

The subjective assessments of children's inattentive, hyperactive, and impulsive behaviours are shaped within diverse cultural and social contexts. When directly comparing ADHD symptoms among groups of children from different cultural backgrounds, it becomes challenging to discern whether differences arise from variations in the children themselves, the raters' differences, or a combination of both. The comparability of ADHD prevalence across various cultures and societies might be compromised by divergent perceptions of ADHD-like behaviours.

Cross-cultural research has aimed to investigate the cultural factors influencing judgments of children's ADHD-like behaviours. Mann et al. (1992) examined the impact of cultural differences on ADHD prevalence by enlisting mental health professionals from China, Japan, Indonesia, and the United States to rate the intensity of hyperactive and disruptive behaviours displayed by four 8-year-old boys in videotaped recordings. Chinese and Indonesian clinicians assigned significantly higher ratings to hyperactive and disruptive behaviours than their Japanese and American counterparts. This finding indicates substantial variations in perceptions of such behaviours across different countries, even when standardised criteria are employed. However, this study focused solely on healthcare professionals who generally do not directly observe children's behaviour, potentially limiting the generalisability of the results. Nevertheless, the utilisation of standardised videotape vignettes to explore cultural factors in the subjective assessment of children's ADHD behaviours has inspired subsequent cross-cultural investigations.

Building upon Mann's research, Muller et al. (1995) employed videotape vignettes to explore teachers' ratings of disruptive behaviour across five countries: China, Indonesia, Japan, Thailand, and the United States. Using vignettes featuring four 8-year-old boys (two Japanese and two American), the study gathered responses from 130 teachers. Comparative analysis of their ratings revealed that American and Japanese teachers perceived disruptive behaviours as less severe than teachers from China, Indonesia, and Thailand when observing children playing individually. However, the differences in perception were less pronounced when observing children engaged in group activities.

Alban et al. (2002) examined the ratings of ADHD behaviours by teachers and student teachers in Mainland China, Hong Kong, and the United Kingdom. Participants (373 in total) viewed a 28-minute video depicting the behaviours of a 9-year-old white Caucasian boy. They evaluated his behaviours using a checklist covering inattention, hyperactivity, impulsivity, and peer relations. The analysis of teachers' ratings across these cultural settings revealed that Mainland Chinese teachers rated hyperactivity and inattention significantly higher than their Hong Kong and UK counterparts.

Curtis et al. (2006) utilised text vignettes to compare how teachers in the United States and New Zealand perceive interventions for children with ADHD. They presented three text vignettes describing distinct types of ADHD symptoms and assessed teachers' perceptions of four different interventions. The study identified a notable difference in intervention perceptions between American and New Zealand teachers.

Given the crucial role culture plays in diagnosing ADHD in children, quantifying the extent to which rater bias contributes to disparities in ADHD diagnoses is essential. While the cross-

cultural studies mentioned above have illuminated differences in teachers' perceptions of children's ADHD behaviours, limitations in text and video vignettes hinder the comprehensive identification of such perceptions. Additionally, these studies did not thoroughly explore the degree to which rater differences contribute to cross-cultural discrepancies in ADHD diagnoses, limiting their capacity to enhance cross-cultural diagnostic comparisons of children's ADHD.

3.5 Cross-cultural Comparisons of Prevalence

The methodology involving the direct comparison of samples across different cultural contexts for assessing ADHD symptoms has generated considerable controversy. Numerous cross-cultural comparative studies have directly contrasted the ADHD diagnosis outcomes of distinct groups of children, often neglecting to account for rater bias. Recognising these challenges, Veit et al. (2007) devised a two-step approach to minimise methodological biases when comparing Brazilian and German children with ADHD/HD. Initially, they conducted culture-specific assessments aligned with clinical diagnostic criteria to identify children with ADHD in both countries. Brazilian participants were diagnosed based on DSM-IV criteria, while their German counterparts were diagnosed using ICD-10 criteria. Subsequently, parents rated children, both with and without ADHD, using the Children Behavior Checklist (CBCL) to assess psychological issues linked to ADHD/HD. This standardised CBCL scale was employed in both countries.

Comparing parental ratings across the two nations on eight scales (Withdrawal, Somatic Complaints, anxiety/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior), the researchers found that Brazilian children received significantly higher ratings than their German counterparts across all eight CBCL scales. Nonetheless, the two countries exhibited similar CBCL profiles in terms of parents' assessments of behavioural and emotional problems in children with or without ADHD.

However, despite attempts to mitigate methodological biases, this research introduced potential complications. Firstly, the adoption of different diagnostic criteria (DSM-IV and ICD-10) by the two countries to identify ADHD children, though referred to as culture-specific assessments, implies divergent diagnostic criteria. Extensive research has demonstrated varying performance of DSM-IV and ICD-10 criteria in identifying children with ADHD (Doernberg & Hollander, 2016; Lahey et al., 2006). DSM-IV's broader ADHD description contrasts with ICD-10's criteria, and the two frameworks differ in addressing comorbidity. Therefore, comparing populations identified using distinct assessment instruments raises questions. Secondly, detailed information regarding the CBCL instrument's consistency and reliability for culturally diverse samples is absent. While parents from both countries utilised the same CBCL rating scale, its uniform performance across the two countries remains uncertain. These limitations highlight the intricate nuances inherent in cross-cultural comparisons, necessitating a comprehensive evaluation of diagnostic tools' performance and consistent application across diverse cultural settings. Addressing these issues becomes imperative in fostering more equitable and accurate cross-cultural comparisons in the realm of ADHD research.

Another international comparison of children's ADHD behaviours was conducted through a different methodology. Merrell et al. (2013) examined inattentive, hyperactive, and impulsive behaviours among young children across England, Scotland, and Australia. This comprehensive international comparison entailed a substantial sample of 2,500 children per country. Teachers assessed children from the three nations using a rating scale based on DSM-IV diagnostic criteria for ADHD. Before comparing children's ratings across countries, the researchers employed the Rasch measurement model to investigate construct consistency and scale reliability across distinct cultures. This model facilitates mathematical analysis to

determine whether items measure the same construct, creating a unidimensional equal-interval scale for measuring children across cultures. The Rasch model's application can investigate scale performance between various groups of people, such as differing gender and cultural groups.

Merrell et al.'s study harnessed the Rasch Model to craft a unidimensional equal-interval scale, enabling measurement of all three groups of children from diverse cultural backgrounds on the same scale. Biased items across gender and cultural groups were discriminated against, resulting in the creation of an equitable measurement scale. Comparing children from the three countries on this standardised scale revealed higher ratings for Scottish children in inattentive, hyperactive, and impulsive behaviours by their teachers than their counterparts in England and Australia. The study suggests several potential factors contributing to these differences, encompassing variations in ADHD behaviours among Scottish children, teacher overrating, or a blend of both factors.

While this cross-cultural comparison employed a uniform instrument across the three countries and utilised the Rasch Model to establish equal performance, the influence of raters from distinct cultural contexts introduced potential discrepancies in perceptions of children's behaviours. Thus, discerning whether disparities in children's behaviours stem from divergent raters' perspectives, the actual behaviours exhibited, or a combination thereof remains complex. Merrell et al. emphasised the necessity for further research to disentangle these possibilities. They proposed that requesting teachers to rate the same video-recorded behaviours might offer a potential solution for distinguishing differences in teacher perceptions.

This raises the question of how to rectify ratings of children's ADHD behaviours to facilitate equitable cross-cultural comparisons. To achieve this, three primary aspects need addressing. First, the instrument employed to assess children's ADHD behaviours must demonstrate equal performance across cultures. Using the same instrument does not guarantee uniformity in measuring different groups. Items may exhibit differing performance across groups; behaviours considered acceptable by Chinese teachers might be deemed unacceptable by English educators, for instance. A robust analytical procedure to evaluate instrument performance is essential for cross-cultural comparisons.

Secondly, the influence of rater bias must be accounted for in producing equitable comparisons. Measuring and identifying rater bias across cultures can differentiate this influence from children's behavioural ratings, enabling a more equitable international comparison of children's behaviours. Finally, adopting a measurement model that considers not only children's behaviours but also potential influences, such as item difficulty and rater factors, is crucial to adjusting children's behavioural rating scores, thereby minimising these influences to facilitate fair cross-cultural comparisons.

Delving into the cross-cultural comparison of teachers' ADHD behaviour ratings within classroom settings helps to gain a more nuanced understanding of the intricate interplay between raters and culture in shaping ADHD diagnoses. This dynamic is exemplified through a comparison of the United Kingdom and China, two nations with well-established and notably disparate cultures spanning across continents. Beyond cultural distinctions, the intricate tapestry of societal, political, economic, and educational factors further delineates the divergence between the UK and China. These multifaceted influences contribute to varying ADHD prevalence rates, distinct diagnostic approaches, and divergent treatment modalities

within these nations. Such a cross-cultural exploration sheds light on the intricate interplay of contextual factors that underscore the complexity of ADHD diagnosis, driving the need for a comprehensive and culturally sensitive approach to understanding and addressing this condition.

3.6 ADHD in the UK

3.6.1 The History of ADHD in the UK

Although hyperactivity was first mentioned in the British Medical Journal in 1968, it was believed to be a symptom of other underlying conditions rather than a disorder (Graham & Rutter, 1968). At that time, British psychiatrists operated under a different medical model from America. They were reluctant to accept ADHD as a childhood mental disorder. Unlike the American medical model, in which psychiatrists rather than psychologists dominated, UK psychiatrists were less keen to embrace the biomedical model of mental illness than their American counterparts. British psychiatrists believed that childhood psychiatric disorders often had roots in psychosocial stress rather than biological deficits, and British psychiatrists were very averse to medication for treating childhood psychiatric disorders.

ADHD was initially seen as a North American problem rather than a general childhood mental disorder (Smith, 2017). In the 1990s, British psychiatrists began to favour the North American approaches to hyperactivity (Bishop, 2010). However, the prevalence of hyperactive disorder across the UK and America varies. The prevalence in British children is much lower than in American children (Faraone et al., 2003). This difference was ascribed to the fact that the two countries had adopted different diagnostic criteria. The North American physicians had adopted the DSM criteria, while their colleagues in the UK preferred the WHO's International Classification of Disease (ICD), which was more stringent than the DSM. During the 2000s, public awareness of ADHD increased, and the DSM guidelines were adopted to diagnose children with ADHD in the UK. The estimated prevalence of ADHD in UK children under 15 is 3.62% for boys and 0.85 % for girls (Hire et al., 2015).

3.6.2 Diagnostic Criteria and Treatment

Meanwhile, it has become common to use stimulant drugs to treat this disorder in the UK. Methylphenidate and dexamphetamine were the first stimulant medications with proven efficacy in treating core ADHD symptoms (Brown et al., 2005). These compounds are now the standard pharmacological method for treating ADHD symptoms worldwide in adults and children. Research has shown an increase in drug use for treating ADHD in children between 1995 and 2008 in the UK (Beau-Lejdstrom et al., 2016). The prevalence of ADHD medication in children was 1.5 per 10,000 in 1995 but had risen to 50.7 per 10,000 children by 2008. The steepest increase in the prescription rate of boys aged 10 to 14 was found among all age groups till 2015 (Renoux et al., 2016).

The diagnosis, treatment and intervention of ADHD involve not only psychiatrists but also schools and social resources. Therefore, it is not cheap. It is estimated that the NHS has an annual expenditure of approximately 23 million pounds on the specialist assessment of ADHD in England and Wales (Holden et al., 2013). In addition, 14 million pounds is devoted yearly to health, social and educational services for ADHD (Holden et al., 2013). On top of that, the NHS spent approximately 35.5 million pounds on prescriptions for stimulant drugs in 2010 (Holden et al., 2013).

Although ADHD has been affirmed as a medical condition by the National Institute for Health and Clinical Excellence in England and Wales (NICE), NICE is less confident about root causes and treatment. NICE emphasizes that apart from genetic factors, environmental influences, diet, and psychosocial factors are other causes. Besides stimulants, educational, behavioural, psychological and dietary interventions are other optional treatments for ADHD. NICE emphasizes that stimulant drugs should only be used to treat severe ADHD symptoms, while for mild and moderate ADHD, parental training should be the first line of treatment

(NICE, 2018). Although there is a rise in the use of stimulant drugs to treat ADHD in the UK, the UK has one of the lowest rates of pharmacological intervention in the treatment of Children with ADHD. In the UK, only 0.02%-1.3% of Children with ADHD were prescribed medication to treat their ADHD (Murphy et al., 2014). So only a small minority of children diagnosed with ADHD receive pharmacological treatment in the UK.

3.6.3 Teachers' Knowledge of ADHD

In the UK, a survey of 800 primary and secondary school teachers about their knowledge of and attitudes toward ADHD showed that most (72%) teachers believed that ADHD affects school children and impacts their life chances. Many of them are aware of the difficulties that children with ADHD have at school. 68% of teachers thought that children with ADHD performed worse at school and ran the risk of being excluded from school. 59% of teachers were concerned about Children with ADHD being bullied at school. Meanwhile, the majority (69%) said their school has a good structure for supporting children with ADHD. However, 42% of the teachers had not received training to help children with ADHD, and many did not recognise some essential symptoms of ADHD, such as impulsive behaviour (41%) and difficulty with organisation (74%). Some of the teachers held different opinions toward ADHD. 21% of the teachers did not believe that ADHD was a mental health condition, and 23% would not refer a child they suspected of having ADHD to a healthcare professional (Savanta ComRes Polls, 2017).

3.6.4 British Culture and Teacher's Role in the Classroom

The British values encompass democracy, the rule of law, individual liberty, mutual respect and tolerance of those with different beliefs (https://www.stepbystepschool.org.uk/smsc/).Under the influence of these values, British education values creativity, independent thinking and individual development.

First, the relationship between teachers and students is equal. As Durkin (2004) said "A pattern of the social constructivist learning environment, where a teacher acts as a guide and facilitator

considering the students as their equals rather than being a guru-like role of absolute authority and knowledge". Hofstede (2012) also said that "*the entire system is based on students' well-developed need for independence and the quality to an extent is determined by the excellence of students*". Being respected as an individual with a home language, culture, life experience, and intellect are values espoused by British educational culture. Therefore, children are expected to show respect not only to teachers but also to each other. Meanwhile, questioning lies at the heart of teaching and learning in British education (Primary National Strategy, 2004). Children are encouraged to question teachers and each other and to express their thoughts and opinions without being judged right or wrong.

Second, the teaching approach is more friendly and personal. British teachers give relatively more freedom to students in their learning. Classroom activities are diverse. In addition to the lecture, group working, watching videos, browsing the internet, gaming, quizzes and discussion are commonplace. Children normally work in groups rather than individually.

3.7 ADHD in China

3.7.1 The History of ADHD in China

Chinese psychiatry has been influenced by Chinese culture and historical factors, such as the prevailing policy climate. In general, Chinese psychiatrists used to reject Western psychiatric concepts due to political and historical reasons, especially during Maoist China (Huang, 2014). However, the situation has changed since the late 1970s, after Deng Xiaoping regained power and launched economic reforms to create a market economy. In the 1980s, connections between China and Western were re-established, and since then, Western ideas have significantly influenced Chinese culture and society. At this time, ADHD as an American psychological concept was introduced in China. Meanwhile, hyperactive behaviours have been recognized by Chinese psychiatrists (Yu-Cun, Yu-Feng & Xiao-Ling, 1985). Since then, many studies have been conducted to investigate this disorder and related treatments in China.

3.7.2 Diagnostic Criteria and Treatment of ADHD

In 1979, the first edition of the Chinese Classification and Diagnostic Criteria of Mental Disorder (CCMD) was published. In China, DSM, CCMD, and ICD are widely used diagnostic criteria for assessing Children with ADHD. A systematic review of 67 studies on the prevalence of ADHD in China shows that almost 87% of studies used DSM as their diagnostic criteria, whereas CCMD was only adopted by 12% of studies (Wang et., 2017). Recent research has found ADHD prevalence in China similar to that in the rest of the world. A systematic review of 67 studies which have been conducted over the past 30 years, involving 642,266 children and adolescents in China has shown that the prevalence of ADHD in China is 6.3% (Liu, Xu, Yan & Tong, 2018).

Treating ADHD with medicine is becoming popular in China. In 2003, 90 million RMB of ADHD medicine was sold. However, by 2005 sales had increased to 300 million RMB (Zhang Xu, 2005). Using Chinese traditional medicine to treat Children with ADHD is quite common in China. In Chinese traditional medicine, ADHD is caused by the imbalance of 'Yin Qi' and 'Yang Qi' in organs (Ni et al., 2014). When 'Yang Qi' is overweight than 'Yin Qi' in the heart, liver and kidney, it causes impulsive and inattentive behaviours. Chinese traditional medicine helps the body to maintain the 'Yin' and 'Yang' balance. In China, acupuncture and auricular point treatments are also used for treating ADHD symptoms. A publication has reported that 80.8% of ADHD patients (aged between 6 and 16) have been cured by acupuncture treatment (Zheng Feng & Zheng Xin, 2017). However, they did not define "cure".

3.7.3 Teachers' Knowledge of ADHD

As ADHD has been widely researched in China for decades, Chinese teachers' awareness and knowledge of ADHD have improved. However, there are still misunderstandings about ADHD among teachers. A survey of 149 primary school teachers found that most teachers (96.6%) had heard about ADHD, and 67% of primary teachers knew that ADHD was a mental disability (Liu, 2013). 13% of primary teachers believed that ADHD was a moral deficiency and that children with ADHD were born to be selfish. Moreover, 12% of primary teachers believed that

children with ADHD were just naughty children lacking discipline and education. Although most teachers (84.6%-86.6%) believed that medicine and psychotherapy would help, 45% of teachers believed that children with ADHD needed severe discipline at school, and 58.4% of teachers required parents of children with ADHD to apply strict discipline to their children at home. Moreover, primary teachers' awareness of ADHD symptoms varies. Most of the teachers (61.7%) are aware of inattention symptoms. However, they are not aware of the symptoms of impulsivity. About a quarter of the informants believed that being active, funny and interrupting conversations are symptoms of ADHD.

As a result of misunderstanding ADHD symptoms, some children without ADHD raised concerns about having ADHD. Although most teachers believed that medical treatment was helpful and knew that medicine has side effects, they did not know what side effects to expect. 49% of informants believed that children would be addicted to the drugs, and 16% believed that the drugs would cause infertility. In addition to lacking knowledge of ADHD and related training in how to teach children with ADHD, most teachers had a negative attitude towards children with ADHD in China (Jin et al., 2010). Most of them (96.4%) thought that children with ADHD negatively affect other children's learning progress, and almost 72% of teachers dislike teaching children with ADHD. 4% of teachers would like to dismiss them from schools, and 12.8% believed that children with ADHD did not deserve any positive feedback for their progress because children with ADHD are presumed to set bad examples for other children. These misunderstandings will not make children with ADHD's lives easier at school or home. Teachers who have an ADHD child in their class do not receive support for teaching the child. Moreover, there is no special education needs support for children with ADHD at home or school.

3.7.4 Chinese Culture and the Teacher's Role in the Classroom

Chinese culture is one of the oldest cultures in the world. Despite influence from outside China, it has preserved some unique characteristics. Traditional Confucian philosophy has

tremendously influenced education in China (Ho, 2006). Confucianism values self-improvement, which includes self-discipline, honouring the family, respect for elders, and the hierarchy (Hui et al., 2011).

First, the relationship between teachers and students has been greatly influenced by Confucian values of respect for authority and seniority (Lehman & Tweed, 2002). Due to the traditional Confucian culture, the classroom formality and the teacher-centred lecturing style, the relationship between teachers and students is not that of equals (Xiao, 2006). Chinese teachers are regarded as authoritarian figures who represent the value of Chinese education and hold the goals and accountability of education. Therefore, students are expected to show respect and obey their teachers. Questioning and discussion without teachers' permission are not encouraged. Moreover, problem behaviours are perceived as challenging the teacher's authority (Gu, 2006).

Second, Chinese educational culture is collectivism which emphasizes the goals of the group or society, fosters ingroup belonging, demands cooperation and interdependence, and pursues harmony (Basu-Zharku, 2011). Individual interests and pleasure have to give way to social norms, goals and needs. Therefore, the teaching methods focus on transmitting knowledge instead of cultivating independent thinking (Tan, 2017). Chinese education places a high value on the result of exams. Teachers are under pressure to prepare students to pass exams. Teaching styles are stricter and more repetitive in the Chinese classroom (Di & McEwan, 2017). Chinese teachers focus on passing on knowledge to students and ensuring they can get the right answers to exam questions rather than encouraging students to explore and think independently. Francis McGonigal (2002) from Birmingham commented, *"Chinese methods encouraged learning by rote, rather than independent thinking"*. Therefore, in Chinese primary classrooms, most teaching activities are lecturing and practising exam questions. Group working, gaming, watching videos, and discussing, are not common activities in Chinese primary classrooms. Therefore, children must stay seated, be quiet, focus, and obey the teacher's instructions.

In Chinese classrooms, students tend to work individually. This is not only due to cultural factors but also to class sizes. Oversized classes are still a common phenomenon in China. The average class size in primary schools is 37.5, whereas the average class size in UK primary schools is 25. (OECD, 2019). However, class size can stretch to 113 students per class due to the lack of teachers and classrooms in densely populated and undeveloped provinces (Yang, 2018). Therefore, under pressure to deliver knowledge and manage oversized classes, Chinese teachers have adopted strict and repetitive teaching methods. From a Western perspective, "*the Chinese teachers are always explainers, and the Chinese students are just listeners and note takers*" (Huang & Brown, 2009, p. 649).

3.8 Conclusion

In conclusion, there are differences in the diagnosis, treatment, and history of ADHD between the UK and China. Moreover, the two countries represent two extraordinarily distinct cultures, and cultural factors play an important role in the diagnosis of ADHD. As one of the valuable informants for ADHD diagnosis, teachers' judgement is based on their knowledge and beliefs. Their beliefs about the cause and their knowledge of the disorder's manifestations and impairments may impact whether they advocate for the referral of children with suspected ADHD (Hillman, 2011). Teachers' beliefs and knowledge are "*rooted in, and constrained by, the culture of the society in which the teachers are living and working, in the culture of the education systems and traditions of the society and in their own experience as school students, teacher education students and members of school communities*" (Perry et al., 2006, 446). Hence, it is crucial to employ a suitable method for identifying and measuring teachers' perspectives on children's ADHD symptoms in diverse cultural contexts to ensure fair comparisons among distinct cultural groups.

Chapter 4 A method Review of Identifying Raters' Bias in ADHD Diagnosis

4.1 Introduction

The presence of rater bias can have a significant impact on cross-cultural comparisons of children's ADHD prevalence. Rater bias refers to the subjective judgments and personal beliefs that raters bring to the assessment process, influencing their perception and rating of ADHD behaviours (Hoyt, 2000). In a cross-cultural context, where diverse cultural norms, values, and expectations exist, rater bias becomes even more pronounced. This is because raters may interpret behaviours differently based on their cultural background and beliefs, leading to variations in the identification and rating of ADHD symptoms across cultures. Consequently, this bias can result in overdiagnosis or underdiagnosis of ADHD in specific cultural groups, leading to inaccurate prevalence rates and disparities in access to appropriate interventions and support.

To investigate rater bias in cross-cultural contexts, numerous studies have been conducted, involving raters from different cultural backgrounds rating the same target's ADHD behaviours. These studies have employed various methods to present target behaviours for raters to evaluate, including written vignettes, videotapes, and direct observation. This chapter provides an overview of the different methods utilis ed in previous studies and discusses their advantages and limitations. Additionally, this chapter introduces a novel method, cartoon animation, which has yet to be extensively used in this field. While cartoon animation has been widely employed in other studies, the researcher believes that it possesses unique strengths for the purpose of this study.

4.2 Method of Written Vignette

Written vignettes are commonly employed to investigate teachers' recognition and perceptions of ADHD cases (Moldavsky et al., 2012; Ohan et al., 2011). Curtis et al. (2006) employed

textual vignettes to discern disparities in teacher perceptions of interventions for children with ADHD in the United States and New Zealand. They crafted three text vignettes, each depicting distinct types of ADHD symptoms. These vignettes detailed a child's behavioural symptoms in alignment with the corresponding ADHD subtypes. See *Table 4* for vignette example 1.

Table 4 Written vignette example 1

ADHD-Predominantly Inattentive Type
Jonathan/Jane is a 9-year-old student with a long history of being easily distracted by extraneous stimuli, has problems keeping his/her attention focused, fails to pay attention to details, and makes careless mistakes in his/her schoolwork…

Note. Curtis et al. (2006)

Another study utilized textual vignettes to explore the influence of ADHD labels on teachers' and education students' perceptions of children's behaviours (Ohan et al., 2011). Four vignettes were created, each presenting a behavioural profile describing a child's ADHD-related behaviours. Two of these vignettes included a statement indicating the child's ADHD diagnosis, while the other two did not contain such a statement. Participating teachers and education students read the vignettes and subsequently completed questionnaires gauging their perceptions of ADHD symptoms. See *Table 5* for vignette example 2.

Table 5 Written vignette example 2

Vignette 1
Eric is a 9-year-old boy. Eric's teacher describes him as always moving, from squirming in his seat to wandering around the classroom, chattering endlessly instead of doing his work. His teacher says that Eric doesn't do what she asks him to do, such as cleaning his desk, despite her constant instructions…

Note. (Ohan et al., 2011)

The utilisation of written vignettes as rating objects for investigating rater opinions and attitudes offers several compelling advantages. Firstly, these vignettes provide standardised scenarios with controlled content, ensuring a consistent presentation of ADHD behaviours across different raters. This standardisation minimises variability in the ratings, thereby enhancing the reliability of the study findings. Additionally, written vignettes are easy and cost-effective to produce, making them a practical choice for research purposes. Moreover, they eliminate potential biases associated with live observations, as raters cannot be influenced by situational factors. This heightened objectivity contributes to the validity of the ratings. Furthermore, written vignettes afford researchers the flexibility to explore specific aspects of ADHD behaviours in a focused manner, enabling a detailed examination of rater opinions and attitudes toward different symptoms. Overall, the use of written vignettes as rating objects presents a robust and efficient approach for investigating people's opinions and attitudes in ADHD assessments.

However, text vignettes have their limitations. Presenting a behaviour pattern with actual behaviours is often more direct and comprehensible than using words. With 18 ADHD symptoms and a broad array of behaviours and contexts associated with these symptoms, describing them solely through text can be challenging. For instance, consider a symptom such as difficulty sustaining attention in tasks and play activities. The contexts encompass classroom lectures, conversations, extended reading sessions, exams, playing with toys, and more. The manifestation of attention issues can vary from becoming unfocused on a task to changing conversation topics before others complete their points, or rapidly losing interest in a toy. Text vignettes struggle to encompass the necessary information and nuanced details about the target child's behaviours in a concise format.

Furthermore, when utilising standardised text vignettes in a study, factors such as vignette length, the number of described behaviours, and the reading complexity of each vignette need to be standardised. This often leaves little room for including comprehensive details. This explains why the vignettes in the studies only outline the target child's symptoms without providing intricate contexts. Moreover, for cross-cultural studies conducted in different languages, translating text vignettes is necessary. Yet, ensuring the translated vignettes retain equivalence and standardisation across languages can be a contentious issue.

Ultimately, individuals form opinions and judgments about ADHD behaviours based on what they have observed, not solely from what they have read. The descriptions contained in text vignettes already represent the opinions and judgments of a child exhibiting specific ADHD symptoms. Therefore, real-life perceptions or acceptability of children's ADHD behaviours may diverge from those derived from reading these vignettes. Due to the reasons discussed above, text vignettes do not adequately fulfil the requirements of studies seeking to compare teachers' perceptions or attitudes towards ADHD symptoms across various cultures and languages.

4.3 Method of Video Tape

Cross-cultural studies have extensively employed standardised videotape vignettes to investigate the varied perceptions of teachers regarding ADHD symptoms (Moldavsky et al., 2012; Ohan et al., 2011). Du et al. (2003) employed videotapes featuring two British children, one boy and one girl, to examine teachers' ADHD symptom ratings across Guangzhou (Mainland China), Hong Kong, and Leeds. These children, both clinically diagnosed with ADHD, were six years old and were recorded in solo and group play activities. Each video, edited to a 20-minute duration (10 minutes solo, 10 minutes group), was assessed by teachers from the mentioned regions against ADHD symptoms.

Similarly, another group of researchers (Alban-Metcalfe et al., 2002) conducted a parallel investigation to explore differences in ADHD ratings between teachers and student teachers across three cultural settings: Guangzhou, Hong Kong, and Leeds. They utilised a videotape vignette from Dyson's (1998) ADHD research, containing 28 minutes of footage extracted from 9 hours of observation on a 9-year-old Caucasian boy with ADHD. Educators from the three regions watched the video and rated the child's behaviours using the SNAP Inventory Scales for Hyperactivity, Inattention, Impulsivity, and Peer Relations.

These studies indicate that Videotape vignettes provide a more extensive collection of information and contextual details compared to text vignettes. Vignettes captured on tape not only portray ADHD-like behaviours but also the settings and circumstances wherein these behaviours transpire. Thus, they afford diverse behaviours in varied contexts, simulating real-world child observation for participants. Furthermore, videotapes transcend language barriers often encountered in cross-cultural studies involving distinct languages, as subtitles can be incorporated to aid comprehension.

However, the use of videotape vignettes does come with limitations. Crafting standardised vignettes is both costly and time intensive. Engaging children and schools requires obtaining consent forms from parents, teachers, and educational institutions. Access to recording equipment and ensuring minimal disruption to children's learning routines further complicates the process. Recording, often requiring hours to yield mere minutes of usable vignettes, demands proficiency in recording and editing. Moreover, the observer effect, wherein children may alter behaviour due to the presence of recording equipment, can inadvertently skew results.

In addition to the practical challenges, it's important to note that even though standardised videotape vignettes are widely used, they come with their own set of limitations. These limitations stem from the need for these vignettes to capture a comprehensive spectrum of behaviours and contextual situations. This requirement necessitates the involvement of professionals who are knowledgeable and experienced in recognising and understanding ADHD-related behaviours. Additionally, these vignettes must limit the reliance on language, as the interpretation should not depend on language comprehension. Striking a balance between the length and complexity of the vignettes is essential to prevent participant fatigue and ensure that individuals engaging with the material can maintain their attention throughout. This consideration is particularly important because the accuracy of the subsequent rating forms hinges on participants' sustained engagement and their ability to provide precise judgments based on what they've observed.

Despite adhering to standards, videotape vignettes can still introduce bias. Ethnicity and gender can influence behavioural judgments. De Ramirez and Shapiro (2005) noted that Hispanic teachers rated ADHD behaviours in Hispanic students more problematically than their White counterparts. Such disparities in standards or tolerance of behaviours based on ethnicity have been observed elsewhere (Lambert et al., 2001). Ethnicity and gender of children depicted in vignettes can thus confound ratings, introducing potential bias.

Moreover, the inherent ethnicity and gender of children in these vignettes can trigger assumptions or stereotypes among raters. For instance, stereotypes may lead to perceptions of Asian children as quiet and Western children as more energetic. Consequently, differences observed in such comparison studies could reflect the raters' perceptions toward children of

specific cultural backgrounds and may not extend to other ethnic and gender groups. This ethnic and gender dynamic within the vignettes may consequently skew ratings.

Lastly, as previously discussed in the chapter on Understanding ADHD Diagnosis and the Role of Culture and Raters, divergences in education systems between China and the UK, encompassing teaching styles, classroom setups, and activities, could render it challenging for teachers from one country to accurately rate behaviours of children from the other's educational context. The need to limit the cultural influence and potential confusion highlights the necessity for teachers from both regions to assess children's ADHD behaviours in a neutral classroom environment devoid of specific cultural influence. Therefore, the application of videotape vignettes featuring Chinese or English children's behaviours does not align with the research objective.

4.4 Method of Observation

Apart from written and videotape vignettes, alternative studies have employed objective observation to compare with teachers' ratings, aiming to scrutinise cultural disparities in these assessments. Sonuga et al. (1993) endeavoured to investigate inter-ethnic bias in teachers' ratings of childhood activity and inattention. They compared teachers' ratings from standard questionnaires and structured interviews with observational measures for the same cohort of children. Interestingly, while teachers rated Asian children as hyperactive as English children, the observational results demonstrated that Asian children exhibited less hyperactivity than their English peers. This finding suggested a tendency for teachers to overestimate Asian children's hyperactivity, highlighting the potential for bias in teachers' ratings within distinct cultural contexts. This approach raises the possibility that incorporating more objective observational methods alongside subjective ratings could unveil underlying biases in teachers' evaluations across diverse cultural settings.

Weisz et al. (1995) extended this exploration by comparing teachers' reports with direct observations of children's behaviours across two countries (Thailand and the US). The study revealed a stark contrast between Thai and American teachers' perceptions of problem behaviours in their respective students. Thai teachers reported higher levels of problem behaviours in Thai children compared to American teachers' evaluations of American children. However, direct observations painted a different picture, indicating that American children exhibited more problem behaviours than Thai children. This discrepancy underscored the variability in how teachers from different cultures perceived and rated children's behaviours, even when confronted with the same set of actions. Notably, behaviours that Thai teachers might consider minor or typical were rated as serious problems by American teachers, demonstrating the significant impact of cultural context on the interpretation of behaviours.

Similarly, Puig et al. (1999) undertook a comparative study involving teacher reports and direct observations of Jamaican and African American children. Their findings indicated a marked contrast in the perceptions of American teachers and Jamaican teachers regarding problem behaviours exhibited by their students. While American teachers attributed significantly higher total problem behaviours to American children, Jamaican teachers rated their children lower in this regard. However, direct observations of the same group of children revealed a different outcome, with observers reporting higher levels of problem behaviours in Jamaican children compared to their American counterparts.

Notably, Epstein et al. (2005) employed a similar approach to assess the accuracy of teachers' ratings by comparing them to direct classroom observation. This comparison arises from consistent differences in how teachers rate attention-deficit/hyperactivity disorder (ADHD) in

students of various ethnic backgrounds. These disparities prompt inquiry into whether they stem from teacher bias or mirror the genuine behaviour of students within the classroom. Consequently, the distinctions between the evaluations given to Caucasian and African American elementary school children, both in teacher assessments and observed classroom conduct, were examined. The study used complex models to explore the interrelationship between these factors. The results suggested that teachers didn't show bias when rating African American kids with diagnosed ADHD. Similarly, Hosterman et al. (2008) discovered that teachers' ratings of ethnic minority students aligned more closely with direct observation data, enhancing the credibility of teachers' assessments in capturing true behavioural levels.

Nevertheless, while these prior studies utilised direct observation as an objective measure to examine cultural and ethnic factors in teachers' ADHD behaviour ratings, various challenges are associated with this approach. First, the underlying assumption of objectivity in classroom observation warrants scrutiny. The notable discrepancy between teachers' ratings and behavioural observations, as evidenced by low correlations (Epstein et al., 2005), suggests that teachers and observers might not be observing the same behaviours. Second, teachers' ratings inherently encompass observations over a minimum timescale of six months, whereas practical limitations prevent classroom observations from spanning such a prolonged duration. Consequently, certain significant behaviours that transpire beyond the observation period might remain unrecorded. Third, cross-cultural research introduces logistical hurdles to classroom observation. Conducting such observations across diverse countries necessitates assembling a team of trained observers, an endeavour that is both time-consuming and financially demanding.

Furthermore, objective observation in cross-cultural research presents substantial challenges for the observers themselves. Proficiency in the languages utilised in both cultural contexts, coupled with familiarity with distinct conventions, gestures, slang expressions, school routines, and classroom activities in each culture, becomes imperative. Given the scope of my research involving a substantial sample of children and schools from both countries, meeting these requirements within the available resources presents practical difficulties. Hence, objective observation, while valuable in theory, is not a feasible option for my research.

4.5 Method of Cartoon Animation

The original meaning of "cartoon" derives from the Italian word 'cartone,' signifying 'big paper,' which referred to a full-size drawing made on paper as a preparatory study for artwork (Benton, 2009). This term was initially associated with fresco painting, a technique employed in the sixteenth century for applying pigment to wet lime plaster (Benton, 2009). In the 1800s, the British humour magazine Punch utilised cartoons to humorously satirise the Parliamentary Committee (Kemnitz, 1973). Over time, the meaning of cartoons has evolved to represent "a humorous drawing." After centuries of transformation, this word and technique now hold different meanings and applications in the modern world.

In modern times, a "cartoon" is typically defined as a drawing, often found in newspapers or magazines, designed to convey humour, offer political critique, or serve as the basis for a humorous TV program, short film, or other media. These cartoons frequently employ drawn characters and images rather than real ones ("Cartoon," 2019). Animated cartoons of this kind are immensely popular among children and young people, exerting a significant influence on them. These animations often feature anthropomorphised animals, superheroes, child protagonists, and similar themes. One of the earliest and most renowned cartoons, and the first to incorporate synchronised sound, is Walt Disney's "Steamboat Willie," released in 1928,

which marked the debut of the beloved character Mickey Mouse (Batchelor, 2009). This cartoon laid the foundation for the global animation empire that exists today.

One of the key reasons behind the immense success of these cartoons is their simplicity, entertainment value, humour, and universality, as they transcend language barriers. As a result, people from diverse cultures worldwide can understand and enjoy these stories, making cartoons a truly global medium. In the contemporary world, cartoons have emerged as potent and effective tools for communication. Creators use cartoons to convey their perspectives on the world, express opinions and beliefs, entertain, persuade, educate, and even express outrage. Cartoons can be controversial, evoking strong emotional and political reactions, or they can be funny, educational, satirical, and challenging societal norms. Consequently, cartoons find applications in a wide range of fields, including entertainment, advertising, education, and research.

Cartoons set themselves apart from traditional forms of communication, such as text and photos, through their conciseness. They offer a "bigger picture" by conveying a message beyond the immediate image, efficiently packing a wealth of information that can be grasped swiftly. Communicating complex or tedious ideas often requires hundreds of words, whereas a single image in a cartoon can achieve the same goal more effectively. Consequently, cartoons serve as efficient communication tools, making intricate or dull concepts more accessible and encouraging engagement.

In the realm of education, cartoons prove to be valuable resources. Research has demonstrated that children can enhance their language skills through exposure to cartoons (Lodhi et al., 2018; Alexiou, 2019). When animated cartoons are used to illustrate scientific and social topics,

students exhibit significantly better conceptual understanding compared to traditional text and diagram-based approaches. Cartoons substantially improve children's learning abilities and impart lessons on interpersonal issues like family dynamics, and friendships, as well as social and cultural topics (Toledot et al., 2014; Kaptan & İzgi, 2014). Moreover, cartoons have found applications in various assessment methods in schools. For instance, Singaporean schools have employed cartoons for formative assessments to evaluate children's comprehension of biological inheritance (Chin & Teou, 2010), while American schools have used cartoons to assess attitudes toward school (Arlin & Hills, 1974).

Cartoons serve as effective communication tools in cross-cultural contexts. Traditional language-based communication often falls short when bridging gaps between diverse cultures and languages. Social science researchers have uncovered the remarkable effectiveness of well-crafted cartoons, offering a level of engagement and comprehension that few other media can rival in cross-cultural communication.

Animated cartoons have proven valuable in helping individuals from different backgrounds and cultures comprehend complex topics, including health issues. A study comparing the impact of a printed message on polio vaccinations with an equivalent message presented in the form of animated cartoons demonstrated that a well-designed animated cartoon is more effective in delivering a message than written instructional materials (Leiner et al., 2004). Similarly, research has found that cartoons outperform verbal or written instructions in educating patients preparing for colonoscopies (Maxwell et al., 2014).

Furthermore, cartoons are a versatile tool when aiming to elucidate culturally distinctive subjects. For instance, Wushu cartoons have effectively conveyed Wushu culture on an

international scale (Ji, 2010). Their flexibility and ease of production make cartoons a preferred choice over resource-intensive visuals. Stephen et al. (2006) developed a Cartoon-based Hostile Attributional Bias Measurement, which is both culturally sensitive and suitable for assessing urban African American girls' perceptions.

Due to these numerous advantages, animated cartoons were selected as the preferred method to assess teachers' attitudes towards ADHD symptoms across different cultures in this research. There are several reasons behind this choice:

Firstly, cartoons offer a more efficient and illustrative medium compared to written descriptions. They allow for a comprehensive and engaging depiction of ADHD-type behaviours through dynamic visual storytelling. The diagnostic criteria for ADHD symptoms primarily rely on descriptions of specific behavioural patterns and characteristics. These descriptions can often be intricate and challenging to fully grasp through written text alone. Animated cartoons, however, offer a more engaging and illustrative medium for conveying these behaviours. The use of animation enables a dynamic and visually rich representation of ADHD-associated behaviours. In an animated format, these behaviours can be portrayed in a contextualised and narrative-driven manner. Instead of simply reading about these behaviours, teachers can observe the characters' actions, reactions, and interactions with others in various scenarios. This enables a more comprehensive and vivid depiction of ADHD-type behaviours.

Secondly, in cross-cultural research, using animated cartoons to depict ADHD behaviours without language can be advantageous in minimising potential biases that may arise due to language translation. Since cartoons are not reliant on verbal communication, they can effectively convey the intended behaviours and scenarios across different language settings.

This ensures that the assessment maintains its consistency and remains unbiased, as respondents from various cultural and linguistic backgrounds can interpret the animations without the need for translation. This reduces the risk of misinterpretation or miscommunication.

Thirdly, animated cartoons can create standardised contexts that video productions may struggle to achieve, particularly when dealing with teachers from different classroom settings, teaching styles, and activities in China and England. This means that animated cartoons can depict classroom scenarios in a way that is consistent and comparable across different cultural settings. In the study, China and England have distinct educational systems and teaching practices, which can result in varied classroom dynamics and activities. Creating videos that accurately represent both cultural contexts could be challenging, and respondents from one culture may find it difficult to relate to or interpret the scenarios from the other culture accurately. On the other hand, animated cartoons offer the flexibility to construct scenarios that are culturally neutral and universally understandable. The cartoon animations can be designed to depict generic classroom settings and teaching activities that are familiar to both Chinese and English teachers. By avoiding culture-specific references and details, the cartoons ensure that respondents from culturally diverse backgrounds can relate to and understand the scenarios presented. This aspect of standardisation in animated cartoons helps in minimising potential misunderstandings or misinterpretations due to cultural differences, making the assessment more reliable and applicable across different cultural contexts.

Fourthly, the use of animated cartoons offers several practical advantages over video productions. First and foremost, animated cartoons are easier and more cost-effective to create, making them a convenient choice for this research. Unlike video productions that often require

elaborate setups, equipment, and professional filming, creating animated cartoons involves relatively simpler tools and techniques, reducing the overall costs and time investment. Moreover, animated cartoons eliminate the observer effect, which can be a significant concern in research involving live observations or video recordings. The observer effect refers to potential changes in the behaviour or responses of participants when they are aware of being observed. In this context, using animated cartoons instead of live observations or videos ensures that the participants, such as children, do not modify their behaviour or responses due to the presence of observers or cameras. This helps to maintain the authenticity and naturalness of their reactions, enhancing the validity of the study. Additionally, animated cartoons offer greater flexibility in editing and modification during the research process. Unlike video productions that are fixed once recorded, animated cartoons can be easily edited, revised, and adjusted as needed. Researchers can fine-tune the content, pacing, and visual elements of the cartoons to optimise their effectiveness in conveying ADHD behaviours and ensuring cross-cultural relevance. This adaptability allows for a more refined and tailored research approach, increasing the study's precision and reliability.

Lastly, the choice of using animated cartoons offers inherent advantages in terms of engagement and participant involvement in the study. Unlike traditional worded questionnaires that can be perceived as mundane or tedious, animated cartoons are inherently more engaging and captivating. The dynamic and visually appealing nature of cartoons captures the attention of informants, especially teachers, who may otherwise find completing a questionnaire to be a monotonous task. Moreover, animated cartoons create a friendly and approachable atmosphere for the informants. The use of animated characters and storytelling elements fosters a sense of familiarity and comfort, making the research experience more enjoyable for the participants. This friendly and inviting tone helps to establish a positive rapport between the researcher and

the informants, which, in turn, can positively influence their willingness to participate actively in the study. The enjoyment factor associated with animated cartoons further incentivises informants to be more involved in the research process. Participants are more likely to invest time and effort in providing thoughtful and accurate responses when they find the research experience enjoyable and interesting. As a result, the quality and richness of the data collected are likely to improve, contributing to the overall effectiveness of the study.

In conclusion, animated cartoons were deemed the most effective method to portray children's ADHD behaviours in this research, enabling measurement of rater differences in the assessment of children's ADHD behaviours between Chinese and English teachers.

4.6 Conclusion

In conclusion, the review of different methods of portraying children's ADHD behaviours to explore rater bias highlights the advantages and limitations of each approach, with cartoon animations standing out as a promising tool in cross-cultural research.

Using written vignettes to portray children's ADHD behaviours for investigating rater opinions and attitudes in ADHD assessments has advantages. They offer standardised scenarios for consistent presentation of ADHD behaviours, are cost-effective, eliminate observer bias, and allow focused examination of specific symptoms. However, they may have limitations, such as providing limited information compared to visual mediums like cartoons or videos, facing challenges in potential translation issues for cross-cultural studies, and possible discrepancies between perceptions formed from reading vignettes and real-life observations.

Using videotape vignettes offers a comprehensive view of a child's behaviours and surroundings with various contexts, but it has drawbacks. Creating standardised vignettes is costly and time-consuming, requiring school access and consent. The observer effect may influence children's behaviour during recording. Representing a wide range of behaviours in different contexts is challenging. Ethnicity and gender in vignettes can introduce rating bias.

Using direct observation to investigate cultural and ethnic issues in teachers' ratings of children's ADHD behaviours offers the advantage of objectivity and real-time insights. However, it presents challenges, including potential discrepancies between teachers' ratings and observations, limited duration of observations, and impracticality for cross-cultural research due to language and cultural barriers. Given these constraints, alternative methods like animated cartoons may be more suitable for exploring rater biases in rating children's ADHD behaviours across cultures.

Chapter 5 Improving Measurement Reliability and Addressing Rater Bias: An Overview of the Rasch Models

5.1 Introduction

Assessing and comparing ADHD symptoms across different cultural settings is a complex undertaking that requires careful consideration of numerous factors. One crucial factor that can significantly influence the accuracy and validity of such assessments is rater bias. When rating ADHD symptoms, raters, such as teachers, bring their own subjective judgments, beliefs, and cultural influences into the evaluation process, introducing variability in how ADHD symptoms are perceived and assessed. These discrepancies can impede the comparability of ratings across diverse cultures, resulting in inaccurate prevalence rates and limited insights into the cross-cultural manifestation of ADHD. However, while it may be challenging to identify all potential sources of rater bias, I can utilise modern test theory to measure rater leniency or severity.

In modern test theory, statistical models like the Rasch model and the Many-facets Rasch model empower researchers and practitioners to analyse test data, considering the characteristics of the items, the individuals being assessed, and the raters' leniency or severity. By employing these models, researchers can effectively measure rater leniency or severity and make necessary adjustments to enhance the accuracy and comparability of ADHD symptom ratings across diverse cultural settings. These models offer a systematic and rigorous approach to identify, quantify, and account for the influences of rater bias, thereby enabling more reliable and valid cross-cultural comparisons. The advancement of measurement techniques in this regard is crucial for developing a comprehensive understanding of ADHD and its manifestation across diverse cultural contexts, ultimately leading to improved interventions, support, and equity in addressing ADHD-related challenges on a global scale.

This chapter provides an overview of the two models, namely the Rasch model and the Many-facet Rasch model, along with the related statistical indexes. I have adopted these models as indicators in data analysis to investigate the reliability of the measurements I have developed for assessing teachers' leniency towards ADHD symptoms and children's ADHD symptoms in two distinct cultural groups, China and England.

5.2 The Rasch Model

5.2.1 The Basic Principals of the Rasch Model

Measuring raters' leniency or severity towards children's ADHD behaviours of two groups of teachers from different cultural backgrounds need an instrument which performs equally for the two groups of teachers. Merrell et al's study has shown us a scientific way to achieve it: the Rasch model (Rasch, 1993). The Rasch model is a psychometric model which George Rasch developed to analyse categorical data. This Probabilistic model for some intelligence and attainment tests (Rasch, 1993) is based on the following principle:

A person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one. (Rasch, 1960, p.117)

Using this principle, George Rasch formulates a data matrix according to persons' responses to items to produce a unidimensional interval scale: the probability of success. The probability is calculated by the difference between the person's ability and the item's difficulty, see *Equation 1*, P_{ni} is probability of success (X=1 success, X=0 failure) when a person n encounters an item i. B_n is the ability of person n and is D_i the difficulty of item i. The probability of success (P) of an item also follows a 'S' shape which is called the Item Characteristic Curve (ICC) shown in *Figure 1*. The ICC shows that for an item, the very low achiever, the probability of getting

a correct answer on the item is close to 0. For the very high achiever, the probability of getting a correct answer on the item is close to 1. An average achiever's probability of getting a correct answer on the item is 0.5. The item difficulty is defined by the ability of the average person whose probability of success on the item is 0.5.

Equation 1

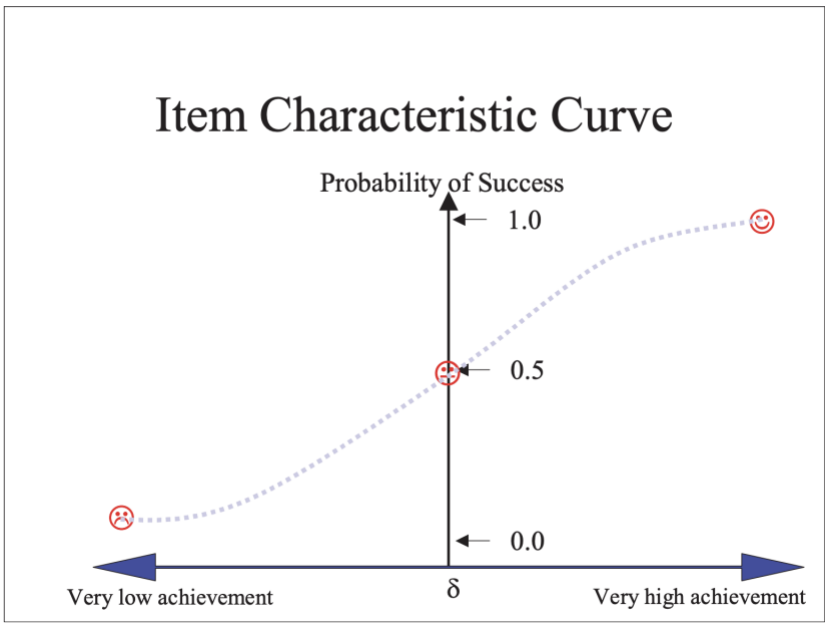
$$P_{ni}(\mathrm{X}=1)=F(B_n-D_i)$$

P_{ni} is probability of success (X=1 success, X=0 failure)

B_n is the ability of person n

D_i is the difficulty of item i

Figure 1 Item Characteristic Curve



Note. (Wu & Adams, 2007)

The difference between a person's ability (θ) and an item's difficulty (δ), which is known as the probability of success, can be transformed into a log odds unit, see *Equation* 2. Therefore, a person's ability which is observed from a raw score (percentage score, what percentage of the correct answers a person gets from the test) and an item difficulty which is observed from the

responses (a percentage, what percentage a person gets a correct answer on this item) are transformed into log odds units (logit). The log odds units produce a unidimensional equal-interval scale for measuring the latent variable. Each item's logit value on the scale shows how difficult this item is, and each person taking the test also has a logit value showing their ability on the same scale. In order to apply the Rasch model to analyse a measurement, the data matrix from the test has to meet two requirements which are *Equation1* and ICC graphical form (Bond & Fox, 2015). This unidimensional equal-interval scale that Rasch Model produces is very important for the study. It means that I can measure two countries' teachers' leniency towards children's ADHD behaviours on one unidimensional equal-interval scale.

Equation 2

$$\log\left(\frac{P}{1-P}\right) = \theta - \delta$$

"Rasch Model states a requirement for the way observations and construct theory combine in a probability model to make measures. There is no other combination of observation and theory that produces sufficient invariance and objectivity in the resultant measures" (Stenner, 2001, p.804). Invariance is a crucial property of scientific measurement. It means that "for any one device, the readings will remain invariant across all suitable contexts, and for any one context, all suitably calibrated devices will yield invariant readings" (Bond & Fox, 2015, p. 84).

Objectivity is another crucial property of a measurement. The Rasch Model has a special property called specific objectivity (Rasch, 1977). The principle of specific objectivity is that a person's trait is independent from the sets of items used to measure it. For example, the difference between two persons' abilities should not be influenced by specific items used. Similarly, the difference between two items only depends on the items' difficulty and is not influenced by which person took the items. Some psychometricians regard this sample-free property of the Rasch model as most important for constructing sound measurements because

statements can be made about relative item difficulties without reference to specific persons. Similarly, statements can be made about the relative proficiencies of people without reference to specific items. These characteristics of the Rasch Model help the study develop instruments that equally measure teachers' leniency and children's behaviours across different cultural backgrounds.

5.2.2 Rasch Model Fit Statistics and Interpretation

The Rasch Model provides a scientific way to investigate the property of invariance of a measurement. I can investigate an instrument's reliability from many parameters to ensure the scales I developed to measure the constructs consistently and reliably before I produce a fair comparison of behaviours between Chinese and English children. The parameters I investigated of the instruments I developed are below:

Category Function: In the Rasch model, the "category function" refers to the relationship between the categories or response options of an ordinal rating scale and the underlying latent trait being measured. The category function assesses how well the different response categories of an item discriminate between individuals with varying levels of the latent trait. In simpler terms, it examines whether the categories on a rating scale are effective in capturing different levels of the trait being measured.

There are a few indices to indicate the function of the category of the scale, such as Fit Statistics, Threshold and Category Function Graph. Fit statistics like INFIT and OUTFIT are used to assess how well the observed responses fit the expected responses according to the Rasch model. Well-fitting categories contribute to a good category function. According to established guidelines (Linacre & Wright, 1994), INFIT and OUTFIT MNSQ values falling within the range of 0.6 to 1.4 are indicative of productive performance for rating scale measurement. The

category function also looks at the thresholds between adjacent categories. It assesses whether there are clear and meaningful distinctions between each category. In a 3-point scale, for instance, there should be distinct thresholds that separate each category. Linacre (1999) recommended a minimum threshold of 1.4 logit for the Andrich threshold, which signifies the minimal advancement in step difficulties between adjacent categories. When analysing the category function, researchers often create a graph that shows the probability of individuals falling into each category at different levels of the latent trait. This graph helps visualize how well the categories are performing. It is called probability curves that shed light on the categorical scale's performance in measuring the sample. The horizontal x-axis signifies the item's level of difficulty, while the vertical y-axis represents the expected likelihood of endorsement for each category when teachers assigned ratings to items. To be considered as functioning reasonably well, it is expected that the highest probability peak for any category should have a minimum threshold value of 0.5.

Person and item statistic summary: they are indications of the reliability of the measurement in Rasch model. It describes the reproducibility of relative measure location which defines the reliability of the measures I get from the scale. There are a few The indications are Person Separation and Item Separation. High reliability of persons means that there is high probability that persons estimated with high measures actually have higher ability than persons estimated with low measures. It tells us whether the instrument discriminates the sample into enough levels for the purpose. High reliability of items means that there is high probability that items estimated with high measures actually have higher difficulty than items estimated with low measures. It tells us whether the sample is big enough to have enough responses to locate the items precisely on the scale.

Wright map: Item and person distributions are presented on a map to show how well the items target the sample persons. This map is called the Wright map. A Wright map makes use of the fact that the difficulty of an item is expressed using the same linear scale (logit scale) that is used to express a person's ability. Therefore, item difficulties and persons' abilities are presented on the same linear scale with an order of difficulty and ability. It helps us visualize where the items are located in comparison with the persons. By this contrast, I can investigate whether items are distributed evenly on the scale in terms of their difficulty values and whether persons are distributed evenly on the scale in terms of their ability. The distributions tell us whether enough items are distributed evenly to discriminate persons' abilities.

Item fit: Item fit investigates whether the item does a good job in discriminating between persons located low and high on the scale and whether the item measures the same latent variable on the scale as other items in the measurement. The Rasch model assumes that a scale measures a single construct, and all the items on the scale should fit the model. Therefore, any item that does not fit well will need further investigation.

Differential item functioning (DIF): DIF investigates an item’s function when it is applied to different groups of persons to identify unfair items (item bias). A scale constructed from an instrument should be valid for all subgroups of respondents. It means that items in the scale have the same relative difficulty across different groups of people, such as different gender groups, different age groups and different country groups. DIF analyses are important in the rating scales development to ensure that the ratings obtained from the scales are unbiased and reflect the same construct for all respondents. In this study, the rating scales measure people across gender and countries. In order to produce invariance scales to measure people across gender and countries, I conduct DIF analysis for the rating scales developed in this study.

In this research, I developed two instruments: Teachers' Rating Scale of Cartoons (TRSC) and Teachers' Rating Scale of Children's Behaviour (TRSCB). The TRSC instrument was developed to identify and measure English and Chinese teachers' leniency of ADHD symptoms according to DSM-V. It helps us to deal with the issue of raters' perceptions that influence the comparison of children's behaviour. The TRSCB instrument was developed to measure English and Chinese children's ADHD symptoms by teachers' reports. The development and evaluation of both instruments followed the Rasch Model analysis to ensure their reliability across the two groups of teachers and children. The fit parameters, as discussed above, were investigated in both instruments with Winsteps software. With the Rasch model, the study can deal with the issue of having instruments that perform equally across different cultural backgrounds.

5.3 Many-facet Rasch Model

5.3.1 The Basic Principals of the Many-facet Rasch Model

The study introduced the Teachers' Rating Scale of Children's Behaviour (TRSCB) to assess children's ADHD behaviours in both China and England. However, as TRSCB is a rater-mediated scale, there can be significant variability among raters when evaluating the observed behaviours of children. Some raters may exhibit more leniency than others, irrespective of their training or the clarity of scoring instructions (Styck et al., 2021). Such differences in raters can potentially introduce bias in participants' trait scoring. To ensure fair comparisons, the study requires a measurement model that considers not only the two traditional parameters of measurement, i.e., the person's ability and item difficulty, but also incorporates the rater's leniency as the third parameter. This model will enable adjustments to the ratings of children's ADHD behaviours, thus mitigating the influence of varying teachers' leniency.

A class of measurement models extends the basic Rasch model by incorporating additional variables or facets, such as raters, scoring criteria, and tasks. Generally, a facet refers to any factor, variable, or component in the measurement process that systematically affects the test score (Wolfe & Dobria, 2008). These measurement models are collectively known as Many-facet Rasch Measurement (MFRM) (Linacre, 1989). They aim to provide a comprehensive analysis of multiple variables (facets) that may impact test or assessment outcomes. The MFRM approach has found increasing applications in diverse fields, including language testing, educational and psychological measurement, health sciences, and others (Bond & Fox, 2007). The MFRM model for polytomous scoring can be written as *Equation 3***.**

Equation 3

$$Log_e\left(\frac{P_{nijk}}{P_{nij(k-1)}}\right) = \theta_n - \beta_i - \alpha_j - \tau_k$$

(Bond & Fox, 2007)

P_{nijk}= probability of examinee receiving a rating of on item from rater

$P_{nij(k-1)}$= probability of examinee receiving a rating of on item from rater

θ_n= ability of person

β_i= difficulty of item

α_j= severity of rater

τ_k= difficulty of receiving a rating of relative to a rating of

The equation expresses the log odds of a person receiving a rating rather than a rating on item by rater. The equation shows that the likelihood of a person achieving a high score rather than a lower score on item depends on the person's ability versus item difficulty, rater severity and the threshold of achieving score. Like the Rasch Scale Model, the MFRM extended the equation by introducing the third parameter, rater severity, for estimating the person's score.

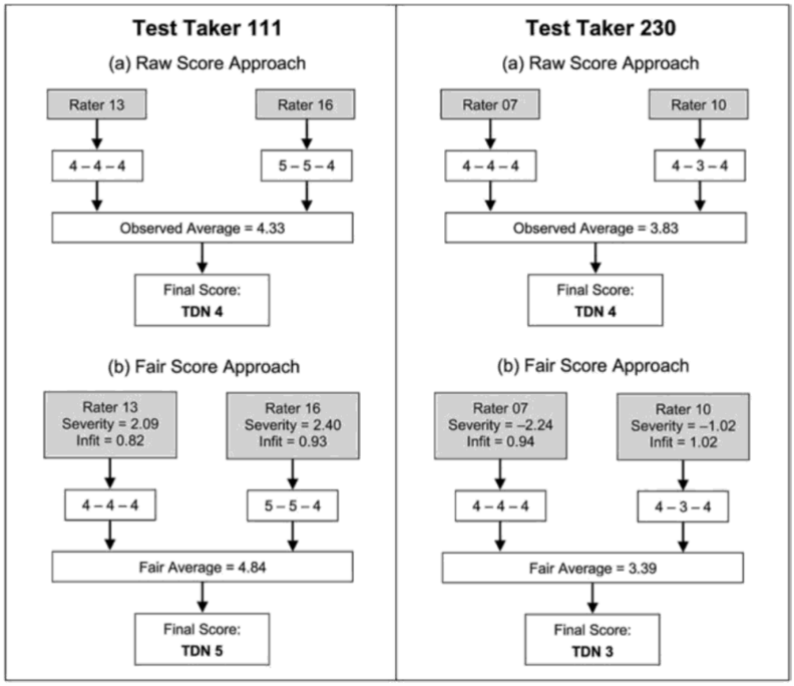
Therefore, this model's essential consequence is that the latent trait is not biased by the rater's differences.

For example, presume person n and person m has the same level of latent. Person n is rated by rater A and person m is rated by rater B. Rater A is more severe than rater B. Therefore, person n will get a systematically lower score than person m. Then it is not fair for person n. However, the MFRM can correct rater's differences and generate a fair estimation of latent. The fair score compensates for severity differences between raters. When computing a fair score, the MFRM model sets all the facets which contribute to the observed score to their mean values, except the latent. In other words, a person's fair score is an expected score that is rated by a hypothetical rater with an average level of serenity. The reference for the average level of severity is the total group of raters in the analysis. So, the fair score of a person's latent is the summation of each score r weighted by the expected probability of achieving a particular score adjusting other facets to average item difficulty and average rater severity.

An example (*Figure 2*) shows the difference between the traditional way and the MFRM way of calculating the final score (Eckes, 2019). There are two test-takers 111 and 230. 111 was rated by two raters 13 and 16. 230 was rated by two raters 07 and 10. With the raw score approach, both test takers get the same final score of 4 by calculating the average of the ratings from different raters. The fair score approach calculates the fair score (fair average) by considering the parameter of the rater's severity. Therefore, the fair score is compensated by the rater's differences. In the example, test taker 111 was rated by severe raters 13 (severity=2.09) and rater 16 (severity=2.40). With the fair score approach, Test taker 111's final score of 4 was upward to 5 to compensate for the severe raters' effect. Test taker 230 was rated by lenient rater 07 (severity=-2.24) and rater 10 (severity=-1.02). With the fair score

approach, test taker 230's final score of 4 was downward to 3 to compensate for the lenient rater effect.

Figure 2 The difference between traditional score and fair score



Note. (Eckes, 2019)

5.4 The Application of Many-Facet Rasch Model in This Cross-cultural Study

However, applying MFRM to correct rater bias requires a special rating or data collection procedure to collect enough ratings to estimate average item difficulty and rater severity. Some studies employed fully crossed, a complete block design to collect data. The fully crossed design means all raters rate all test takers on all traits resulting in no missing data in the data matrix. Although a fully crossed design provides the ideal data matrix for MFRM analysis, it is costly and time-consuming to implement. Some studies employed partially crossed, incomplete block designs to collect data. In a partially crossed design, only a subset of raters rates all the test takers or only a certain subset of raters rates certain traits. Although the partially crossed design is easier to implement, it provides a relatively weak overlap of raters' ratings to estimate other parameters in the analysis.

In my research, the overlap of the rater's rating is necessary for the analysis to adjust the ratings of children's ADHD behaviours according to the teacher's severity. However, collecting overlap rater's rating in cross-cultural research is very challenging. It means that all or some of the English teachers have to fly to China and stay in the classroom to observe and rate the same group of randomly selected Chinese children's behaviours. Meanwhile, all or some of the Chinese teachers have to fly to the UK to implement the same data collection procedure. In practice, this type of research would be very expensive, time-consuming, and hard to implement.

The design of a fully crossed rating or partially crossed rating is to collect overlap ratings to estimate the rater's severity/leniency. As long as I can measure the rater's severity/leniency, I can adjust their ratings according to their severity with MFRM. Rasch analysis software (Winsteps & Facets) allow the users to input known values such as anchored item values (item difficulty) or anchored rater values (rater severity/leniency). So, output (e.g., person values) might be more validly compared across different research contexts. Anchoring parameter's values maintain their invariance, and their values are imported into the whole analysis over the term of a project.

With the Teachers' Rating Scale of Cartoons (TRSC), it measures teachers' leniency towards six symptoms of ADHD (Can't wait, Fidgety, Inattentive, Disorganized, Distracted, Disruptive). The measurements of teachers' leniency will be the anchored value of teachers' leniency in the MFRM analysis of the Teachers' Rating Scale of Children's Behaviour (TRSCB). Therefore, the MFRM analysis generates a fair average score for children's ADHD behaviour by considering the rater's differences. In this way, it can estimate teachers' severity or leniency without requiring Chinese and English teachers to rate students across countries. Moreover, the

MFRM analysis will produce fair scores of children's ADHD behaviours, compensating for the raters' differences across cultures. Then the cross-cultural comparison of children's ADHD behaviours will be a relatively fair comparison by limiting the bias introduced by teachers due to cultural differences.

5.5 Conclusion

In conclusion, this chapter has introduced the Rasch model and the Many-facet Rasch model as essential tools in data analysis for assessing the reliability of measurements in cross-cultural studies. These models were adopted to investigate the consistency of the measurements developed to assess teachers' leniency towards ADHD symptoms and children's ADHD symptoms in two distinct cultural groups, China and England. By utilising these models, the study aimed to address rater bias and ensure fair comparisons of children's ADHD behaviours across different cultural contexts. The application of these models played a crucial role in enhancing the robustness and accuracy of the research findings, ultimately contributing to a deeper understanding of the impact of cultural differences on ADHD assessments.

Chapter 6 Introducing Research Questions

6.1 Introduction

In previous chapters, I explored the various debates surrounding Attention-Deficit/Hyperactivity Disorder (ADHD) and its diagnosis. Through these discussions, four key research questions have emerged. In this chapter, I will present these research questions, provide a background summary, discuss the related research gaps, and outline the proposed methodology to address these questions.

6.2 Question 1

How to measure the leniency of Chinese and English teachers in rating children's ADHD behaviours?

6.2.1 Background Summary

ADHD is a prevalent disorder in children worldwide, with well-established impacts on their lives, families, and societies. Diagnostic criteria and procedures have been developed and refined over time, but the exact cause of the disorder remains unclear. Research indicates that ADHD is a complex condition influenced by genetic factors, brain structure and function, neurotransmitter imbalances, and environmental factors (Thapar et al., 2013; Franke et al., 2018). The interplay of these factors varies among individuals, and there are no definitive laboratory tests or biological markers for diagnosis. Instead, ADHD diagnosis relies primarily on behavioural observations and self-report measures, which can be influenced by cultural, societal, and individual biases.

With the spread of information and increased awareness about mental health conditions, including ADHD, there has been a rise in the recognition and diagnosis of ADHD worldwide. The discrepancy in study findings in this globalisation trend of ADHD indicates a lack of consensus and consistency in the research conducted on ADHD across different regions and

cultures (Rohde et al., 2005; Madsen et al., 2015; Kazda et al., 2021). Different studies may yield conflicting results or varying prevalence rates, leading to discrepancies in the understanding and interpretation of the disorder. The globalisation trend of ADHD highlights the need for a more comprehensive and culturally sensitive approach to studying and understanding the disorder. It calls for cross-cultural collaborations, standardised assessment measures, and consideration of contextual factors to obtain a clearer and more accurate picture of ADHD across different populations. Cross-cultural comparison of ADHD prevalence helps us gain a more comprehensive understanding of the disorder across different populations and identify potential cultural, environmental, and societal factors that may influence its occurrence.

However, there are challenges when comparing children's ADHD prevalence across populations from different countries. First, it is the diagnostic criteria and measurement tools. Different countries may use varying diagnostic criteria and assessment tools for ADHD. This can lead to variations in how ADHD is diagnosed and measured, making it challenging to compare prevalence rates directly. Second, it is the sampling methods and population characteristics: Studies on ADHD prevalence often rely on different sampling methods, such as community-based surveys, clinical assessments, or school-based screenings. The selection of the sample and its representativeness can vary across studies and countries, which may affect the generalizability of the findings. Additionally, differences in population characteristics (e.g., socioeconomic status, cultural factors, access to healthcare) can also impact prevalence rates.

Third, it is cultural and contextual factors. Cultural and contextual factors can influence the perception and manifestation of ADHD symptoms, which can lead to variations in prevalence rates across countries. For example, cultural norms, parenting styles, and educational systems may impact how ADHD symptoms are recognised and reported. These factors need to be

considered when comparing prevalence rates across different cultural contexts. Fourth, it is language and translation issues. Studies conducted in different countries may involve translations of assessment measures, which can introduce challenges in maintaining the validity and reliability of the instruments. Translation errors, cultural nuances, and language structure differences can affect symptom assessment accuracy and lead to discrepancies in prevalence rates.

Last but not least is the underreporting and stigma. ADHD is still stigmatised in some societies, and there may be a reluctance to report or seek diagnosis and treatment (Godfrey et al., 2020; Lebowitz, 2013). This can result in underreporting symptoms and underestimating ADHD prevalence rates, particularly in countries where mental health conditions are highly stigmatised. To address these uncertainties, researchers need to adopt rigorous methodologies, establish standardised diagnostic criteria, and ensure the use of culturally sensitive assessment tools. To achieve that, in the cross-cultural comparison, there is a fundamental issue to address, rater bias. For the diagnosis of children's ADHD, parents and teachers are children's ADHD behavioural observers and raters. As reviewed in the previous chapter, teachers play a crucial role in the diagnosis of ADHD as they have direct and continuous interactions with students in educational settings. Their observations and assessments provide valuable insights into children's behaviours and help identify potential ADHD symptoms. However, the subjective nature of rating ADHD behaviours introduces the potential for rater bias. Rater bias is a significant factor influencing behaviour rating scale scores (Splett et al., 2018). Teachers' perceptions and evaluations of ADHD behaviours can vary due to cultural factors, personal beliefs, and educational contexts. To accurately compare ADHD prevalence across different cultures, it is important to distinguish whether the variations in prevalence rates are due to

actual differences in ADHD symptoms exhibited by children, differences in how teachers perceive and assess those symptoms, or a combination of both factors.

6.2.2 Research Gap

The studies, reviewed in Chapter 3 *Understanding ADHD Diagnosis and the Role of Culture and Raters*, aimed to identify variations in raters' diagnostic practices, specifically focusing on the rating of children's ADHD behaviours. These studies highlighted that teachers from different cultural backgrounds tend to rate these behaviours differently. However, the methods employed in these studies, such as written vignettes, videotapes, and observations, have inherent weaknesses when applied to cross-cultural comparisons. Additionally, these studies did not develop a reliable and valid instrument to effectively quantify rater bias in a cross-cultural context. As a result, there remains a significant gap in accurately and reliably identifying and measuring rater bias across diverse cultural backgrounds.

To address this gap, one of the primary objectives of this study is to develop an instrument that can measure rater bias exhibited by Chinese and English teachers when rating children's ADHD behaviours. In order to achieve this objective, my first research question focuses on understanding how to measure the leniency of Chinese and English teachers in rating these behaviours.

Rater leniency is a commonly observed bias in behavioural ratings, where raters consistently assign either higher or lower ratings compared to their peers, regardless of the actual level of observed behaviours (Dobria, 2011; Erguvan & Aksu Dunya, 2020). This leniency can manifest as a general tendency to rate individuals more positively or negatively than their behaviour warrants. It is influenced by various factors, including personal biases, grading

patterns, and individual differences in rating standards. Cultural biases can also contribute to disparities in rating ADHD symptoms across different cultural groups.

In this study, I aim to measure the leniency exhibited by teachers, as it significantly varies among raters from different cultural backgrounds. One reason for this variation is the nature of children's ADHD rating scales. These scales require teachers to recall a series of ADHD behaviours displayed by the child over a six-month period. Research suggests that teachers are more likely to rely on their general impression of the child's behaviours rather than recalling specific behavioural details (Murphy & Balzer, 1986). Therefore, their overall impression and acceptance of the child's ADHD behaviours play a crucial role in assigning high or low scores.

Moreover, teachers are asked to rate the frequency of the recalled behaviours on a Likert scale, ranging from "never" or "not at all" to "very often" or "severely." While teachers may have a general sense of whether behaviours occur more or less frequently than they prefer, they might have a less defined understanding of what is considered an acceptable level of frequency overall (J. Anthony et al., 2021). Consequently, teacher leniency becomes a significant factor in assigning a low or high frequency to the child's behaviours. As discussed in Chapter 2 about the diagnostic development and criteria evolution, the 18 core ADHD symptoms represent not only challenging behaviours exhibited by children with ADHD but also problematic behaviours that teachers have to manage daily in their classrooms. These behaviours can have an impact on teachers themselves. Their cultural and social values influence how they perceive and interpret these behaviours in their classrooms. Their individual opinions, feelings, and acceptance of these behaviours collectively contribute to their leniency when rating the frequency of the child's recalled behaviours.

Overall, by understanding and measuring rater leniency in the rating of children's ADHD behaviours, this study aims to enhance cross-cultural comparisons and develop a more comprehensive understanding of cultural factors' role in the assessment process.

6.2.3 Research Method

In this study, I have implemented two distinct approaches, cartoon animation and the Rasch Model analysis, to capture and assess the presence of teachers' leniency towards children's ADHD behaviours across different cultural groups. In the previous chapter 4 A Method Review of Identifying Rater Bias in ADHD diagnosis, we addressed the limitations associated with traditional methods like writing vignettes, videotapes, and observation in investigating rater bias across cultures. In our study, we have adopted an innovative approach, cartoon animations, as our primary measurement approach. The utilisation of cartoon animations in this cross-cultural study offers several noteworthy advantages.

Cartoon animations provide a visually appealing and universally understandable medium for presenting children's ADHD behaviours. They effectively engage participants and convey information without language barriers, making them accessible across cultures. Additionally, cartoon animations offer cost-effectiveness and standardisation, making them suitable for large-scale studies. By leveraging these advantages, researchers can enhance the accuracy and cross-cultural validity of assessments regarding the prevalence of children's ADHD.

To implement this approach, I developed six episodes of cartoons depicting a series of children's ADHD behaviours corresponding to six symptoms: "Can't wait," "Fidgety," "Inattentive," "Distractive," "Disorganised," and "Disruptive." Teachers from both China and England were asked to watch these cartoon animations and rate the behaviours exhibited by

the cartoon characters based on their acceptance. This instrument, named the Teachers' Rating Scale of Cartoons (TRSC), was used to measure teachers' leniency.

In addition to cartoon animations, I employed the Rasch Model to analyse the data collected using the TRSC instrument. The Rasch model is a fundamental measurement model in psychometrics that provides a framework for analysing the properties of rating scales and assessing the quality of individual items. By applying the Rasch model, researchers can examine the extent to which rating scales can discriminate between different levels of ADHD symptoms and ensure that the scale items function consistently across different cultural groups. This approach enables the identification of biased items or rating tendencies that may be specific to certain cultural contexts.

Furthermore, the Rasch model allows for the transformation of ordinal rating data into interval-level measures, establishing a common metric that can be used for cross-cultural comparisons. This feature of the Rasch model provides a psychometric framework for measuring teachers' leniency across China and England and facilitates the comparison of values on a common metric. Moreover, the values of teachers' leniency measured through the TRSC instrument can serve as anchored values for adjusting ratings from another instrument (TRSCB) that the study employed in the second research question.

6.3 Question 2

How to make appropriate adjustments to ratings of children's ADHD behaviours to improve fairness in cross-cultural comparison?

6.3.1 Background Summary

In Chapter 3, I reviewed cross-cultural comparison studies that aimed to compare children's ADHD behaviours across different countries. One notable study by Veit et al. (2007) focused

on comparing Brazilian and German children with ADHD/HD. To minimise methodological influences, the study employed a two-step methodology. In the first step, culture-specific assessments were conducted using DSM-IV criteria for Brazil and ICD-10 criteria for Germany to identify children with ADHD. In the second step, parents provided ratings of their children's behavioural and emotional problems using the Children Behaviour Checklist (CBCL) in both countries.

However, there were two key weaknesses identified in this study. Firstly, the use of different diagnostic criteria (DSM-IV for Brazil and ICD-10 for Germany) introduced the potential for bias and raised concerns about the comparability of the identified populations. This variation in diagnostic criteria may have influenced the results and limited the validity of the cross-cultural comparison. Secondly, the study lacked detailed information on the consistency and reliability of the CBCL rating scale across culturally diverse samples. It remains uncertain whether the rating scale performed equally well in both countries. These weaknesses indicate potential limitations in the study's methodology, which may impact the generalizability and reliability of the findings.

To address these limitations, future research in ADHD studies should consider adopting standardised diagnostic criteria and ensuring the consistency and reliability of assessment instruments across different cultural contexts. By employing consistent diagnostic criteria and reliable rating scales, researchers can enhance the robustness and validity of cross-cultural comparisons in ADHD studies.

Merrell et al. (2013) conducted an insightful cross-cultural comparison study that explored the ADHD behaviours of children across three countries: England, Scotland, and Australia. The

researchers employed a rating scale based on DSM-IV diagnostic criteria. They also employed the Rasch Model to ensure instrument consistency and establish a unidimensional equal-interval scale for effective cross-cultural comparisons. Interestingly, the study revealed that Scottish children received higher ratings for ADHD behaviours compared to their counterparts from England and Australia.

However, Merrell and colleagues acknowledged the influence of raters' perceptions, which may vary across different cultural backgrounds, as this can impact the comparison of children's behaviours. The study emphasised the need to discern between raters' subjective perceptions and actual differences in children's behaviours, recognising the potential for cultural biases to affect ratings. This study highlights the importance of taking into account the cultural context and individual perspectives when interpreting and comparing ADHD behaviours across diverse populations.

6.3.2 Research Gap

The second research question focuses on addressing the issue of making fair adjustments to ratings of children's ADHD behaviours by considering teachers' leniency. This adjustment is crucial for improving the fairness of cross-cultural comparisons. However, several challenges need to be overcome to accomplish this objective.

The first challenge involves the development of an instrument that effectively measures children's ADHD behaviours based on teachers' ratings in both China and England. This instrument must demonstrate consistency and reliability across samples from both countries to ensure accurate and comparable results. Developing such an instrument requires careful attention to cultural nuances and ensuring its validity in capturing ADHD behaviours in diverse cultural contexts.

The second challenge lies in establishing a measurement model that incorporates teachers' leniency or severity as a facet in the calculation of the final behavioural score. This model should effectively account for the variations in raters' tendencies to assign higher or lower ratings compared to their peers, regardless of the actual levels of observed behaviours. By incorporating this facet into the measurement model, the impact of raters' leniency can be appropriately considered, leading to fairer and more reliable cross-cultural comparisons.

Addressing these challenges will pave the way for a more comprehensive understanding of ADHD behaviours across different cultural backgrounds. It will also contribute to the development of robust assessment tools and measurement models that can effectively account for the influences of raters' leniency, ultimately improving the validity and fairness of cross-cultural comparisons in ADHD research.

6.3.3 Research Method

To address the first challenge, I developed the Teacher's Rating Scale of Children's Behaviours (TRSCB) instrument, which adopts the DSM-5 standardised criteria for ADHD diagnosis. This instrument enables consistent measurement of children's ADHD behaviours based on teachers' ratings in China and England. To ensure the instrument's consistency and establish a comparable scale for cross-cultural comparisons, the study applied standardised criteria for both countries and Rasch Model analysis to investigate the consistency of the scale in a cross-cultural context. This analysis also helps create a unidimensional equal-interval scale that allows for relatively accurate assessments across different cultural contexts.

In order to make appropriate adjustments to ratings based on rater leniency, the study adopted the Many-facet Rasch Model and anchored values. The Many-facet Rasch Model considers

three facets: item (behaviour), person (children), and rater (teacher). It examines the interactions between these facets and estimates the relative difficulty of items, the abilities of individuals, and the leniency or severity of raters. Each rater is assigned a leniency or severity parameter, capturing their tendency to assign higher or lower ratings compared to their peers, irrespective of the actual level of observed behaviours. These leniency/severity parameters are estimated based on the ratings provided by teachers for individual behaviours.

However, estimating teachers' leniency/severity directly from their ratings of children's behaviours requires s a fully crossed rating design, wherein each teacher rate all children on all ADHD behaviours. In our cross-cultural study, achieving such a design would require all participating teachers to visit each other's country, staying at schools for over six months to observe the ADHD behaviours of children from the other country. However, such an arrangement is impractical to implement. Since there is no cross-rating between Chinese and English teachers, we are unable to estimate teachers' leniency/severity directly from their ratings of children's behaviours. Consequently, the study measures teachers' leniency/severity separately, utilising a different instrument known as the Teacher's Rating Scale of Cartoons, instead of estimating it directly from their ratings on the Teacher's Rating Scale of Children's Behaviours.

To incorporate these measured values into the Many-facet Rasch Model, the study introduced them as anchored values. Anchored values serve as fixed reference points used to calibrate the rating scale. They facilitate the estimation of other facets (children, behaviours) in the rating scale, ensuring accurate adjustments to ratings of children's ADHD behaviours and enhancing fairness in cross-cultural comparisons. This approach allows us to account for rater leniency

and improve the validity and reliability of cross-cultural comparisons in assessing children's ADHD behaviours.

6.4 Question 3

To what extent does the rater bias affect the ratings of children's ADHD behaviours in England and China?

This research question builds upon previous investigations in two areas: measuring teachers' leniency towards children's ADHD behaviours using cartoon animations (research question 1) and adjusting teachers' ratings of children's ADHD behaviours with their leniency measures (research question 2). Now, in research question 3, the study delved into understanding the significant influence of rater bias on the assessment of children's ADHD behaviours in both countries.

The primary objectives are threefold. First, the study aimed to determine whether there are significant changes in children's ADHD behavioural ratings for each country after adjusting them with teacher leniency measures. Second, the study sought to identify if there are notable differences in ADHD ratings between England and China, and whether these differences exhibit significant changes after adjusting the ratings with teacher leniency measures. Lastly, study aimed to investigate whether gender differences in ADHD ratings have undergone significant changes after adjusting the ratings with teacher leniency measures within each country.

To address these questions comprehensively, the study conducted rigorous statistical analysis using the collected measures and data. By analysing the results, we can gain insights into the impact of rater bias on the assessment of children's ADHD behaviours in England and China.

6.5 Question 4

To what extent are the ratings of ADHD behaviours comparable from teacher to teacher in China?

6.5.1 Background Summary

In Chapter 3, I discussed the influence of culture on the identification and interpretation of ADHD symptoms by parents, teachers, and health professionals. Within a culture, individual knowledge, experience, and attitudes can also impact judgments about ADHD. The diagnosis of children with ADHD heavily relies on raters' judgments of their behaviours. The specific ratings assigned by raters are crucial in determining the presence of core symptoms and making an accurate diagnosis. Hence, it is important to investigate the Inter-rater Agreement of raters.

Inter-rater Agreement is a crucial aspect of rating scale reliability and validity. It assesses the level of consensus or homogeneity among different raters in their ratings. It provides insight into the reliability of the rating scale and its validity in measuring the intended trait. Lack of Agreement among raters indicates potential issues with the design of the scale, as raters may interpret or apply it differently. It can also suggest a need for proper training of raters. These considerations are essential for evaluating the effectiveness of a rating scale.

6.5.2 Research Gap

Previous studies have primarily focused on the inter-rater Agreement between raters from different settings, such as parents and teachers. Limited research has explored the inter-rater agreement among teachers in rating children's ADHD behaviours, with only one study conducted in 1996 involving a small sample of eight teachers (Danforth & DuPaul, 1996). Consequently, the extent of inter-rater agreement among teachers in rating children's ADHD behaviours remains understudied. Understanding this topic is crucial as it sheds light on the

reliability of relying on a single teacher's report of a child's ADHD behaviours in the school setting.

6.5.3 Research Method

To address the research question of how comparable the ratings of ADHD behaviours are among teachers in China, the study conducted an Inter-rater Agreement (IRA) study. This study design involved multiple teachers independently assessing the same set of children's ADHD behaviours. The ratings provided by each teacher were then compared to determine the level of agreement among the raters. Statistical measures, such as Cohen's kappa or intraclass correlation coefficients, were utilised to quantify the agreement level between the raters.

Implementing this study design is relatively complex and cannot be easily conducted through a post-survey. It requires the participation of multiple teachers teaching the same group of children. In England, most primary schools typically have only one teacher for a specific grade level, such as Year 2. Some schools may have a teaching assistant as well. In contrast, in China, it is more common for schools to have multiple teachers teaching the same class, with some schools having five to six teachers teaching different subjects in a class. This allows for the selection of two or more teachers to cross-rate the same group of children using the Teacher's Rating Scale of Children's Behaviour, making it feasible to conduct this type of study in China.

6.6 Conclusion

In conclusion, measuring the leniency of Chinese and English teachers in rating children's ADHD behaviours is crucial for achieving fair and accurate cross-cultural comparisons in ADHD research. This measurement helps address the challenges associated with diagnostic criteria, cultural factors, language translations, which can all contribute to variations in prevalence rates across different populations. By understanding and measuring rater bias, particularly in terms of leniency, researchers can distinguish between actual differences in

ADHD symptoms and variations in how teachers perceive and assess those symptoms. This understanding is essential for developing a comprehensive and culturally sensitive approach to studying and understanding ADHD.

To achieve this, the study employs innovative methods such as cartoon animations and the Rasch Model analysis. Cartoon animations provide a visually appealing and universally understandable medium for presenting ADHD behaviours, while the Rasch Model allows for the transformation of rating data into interval-level measures, facilitating cross-cultural comparisons. Furthermore, adjustments to ratings based on rater leniency can be made using the Many-facet Rasch Model and anchored values, enhancing the fairness and validity of cross-cultural comparisons. By addressing these research gaps and employing robust methodologies, researcher can advance our understanding of ADHD across diverse cultural backgrounds and contribute to the development of standardised assessment tools and measurement models. Ultimately, this research aims to improve diagnostic practices, enhance cross-cultural comparisons, and promote a more comprehensive understanding of cultural factors' role in the assessment process of ADHD.

The next chapter delves deeper into the methodological aspects of the study, specifically focusing on the systematic development of measurement instruments used to assess teachers' leniency towards ADHD symptoms and children's ADHD symptoms in the cross-cultural context of China and the UK. This chapter will detail the step-by-step process of designing, refining, and validating the measurement tools to ensure their reliability and sensitivity in capturing the subtle nuances of ADHD behaviours in both cultural groups.

Chapter 7 Development of Measurement Instruments

7.1 Introduction

This chapter delves into the detailed process of developing the measurement instruments employed in this study. Specifically, the study involved the creation of two measurement instruments: the Teacher's Rating Scale of Cartoons (TRSC) and the Teacher's Rating Scale of Children's Behaviours (TRSCB).

The TRSC entails a comprehensive procedure whereby teachers were presented with a series of cartoon animations portraying various ADHD types of behaviours commonly observed in the classroom. Subsequently, teachers were required to rate these animated behaviours using a carefully constructed questionnaire. Thus, the development of this instrument encompasses two essential components. The first component involves the development of the cartoon animations themselves, ensuring they accurately depict children's behaviours within the classroom setting. The second component entails the formulation and refinement of a comprehensive questionnaire designed to capture teachers' leniency toward these animated behaviours.

In addition to the TRSC, the study also developed the Teacher's Rating Scale of Children's Behaviours (TRSCB). This instrument focuses on teachers' evaluations of children they taught about their behaviours in relation to ADHD symptoms. To develop this rating scale, the study primarily relied on a review of existing ADHD rating scales commonly employed in school screening. This review process allowed us to carefully select and adapt relevant items to create an instrument specifically tailored to this research objectives.

By exploring the development process of these instruments, this chapter provides insights into the careful and detailed approach taken to ensure their validity and reliability. I will discuss the

creation of the cartoon animations, the formulation of the questionnaires, and the selection of items for the rating scale, highlighting the rigorous methods employed in this study.

7.2 The Development of Cartoon Animations

The development of cartoon animation for this study involved a comprehensive process aimed at accurately portraying children's ADHD behaviours in a culturally unbiased manner. Firstly, a multi-method approach, including classroom observation, teacher interviews, and a review of rating scales, was applied to develop an extensive ADHD behavioural bank with detailed contexts. This data collection phase ensured a comprehensive understanding of the behaviours to be depicted. Second, the researcher collaborated with PhD supervisors and school teachers to create a series of carefully crafted cartoon scripts, effectively capturing the nuances of children's ADHD behaviours. Lastly, a team of professional animators was engaged to bring these scripts to life through a series of captivating cartoon animations. This meticulous undertaking ensured the accurate and engaging portrayal of children's ADHD behaviours, thus providing a valuable resource for investigating rater bias and facilitating cross-cultural comparisons in ADHD assessments resource for investigating rater bias and facilitating cross-cultural comparisons in ADHD assessments.

7.2.1 Cartoon Concepts

The initial step in creating animated cartoons involves developing the concepts behind these cartoons. During this phase, it is essential to address some fundamental questions related to the cartoon design project. These questions include determining the purpose of the animated cartoons, identifying the specific information these cartoons need to convey to the audience, defining the target audience, establishing the core theme upon which the animated cartoons will be based, and outlining any intended calls to action. Crafting a comprehensive brief allows for the documentation of answers to these essential questions before embarking on the cartoon creation process.

Furthermore, this brief serves as a guiding document throughout the subsequent stages of cartoon development. It plays a crucial role in keeping both the animation team and the researcher aligned as they progress through the various steps involved in creating these cartoons. Cartoon development is inherently a highly creative endeavour, akin to filmmaking. It encompasses a wide array of tasks, including story writing, scene design, character creation, storyboard development, and camera work, among others. Hence, having a well-defined brief for the cartoon concepts serves as a vital reference point, ensuring that the final work aligns with the project's overarching objectives.

The series of cartoons developed in this study serves as a tool for assessing teachers' attitudes toward ADHD behaviours in classroom settings across two distinct countries, England and China. Therefore, the brief concepts for these cartoons can be summarized as follows:

- The audiences of the cartoons are English and Chinese teachers who teach 6 to 7-year-old children.
- The series of cartoons present the behaviours of ADHD symptoms of children 6 to 7 years old in the classroom setting.
- Each cartoon presents a few (5 to 6) behaviours of one ADHD symptom (i.e. fidgety, inattentive, disruptive).
- Each cartoon also includes one typical (ideal) behaviour, which is opposite to the behaviours of ADHD symptoms.
- The scenes should NOT favour any culture, and the characters should look like 6 to 7-year-old children without gender reference.
- The stories of the cartoons should be simple and clear, with no misunderstanding from either of the cultures.

- The animations of behaviours should be simple and clear, with no misunderstanding from either of the cultures.
- The cartoon watching and rating should be finished within 30 minutes.
- Appropriate background music is preferred, but NO dialogue is involved.

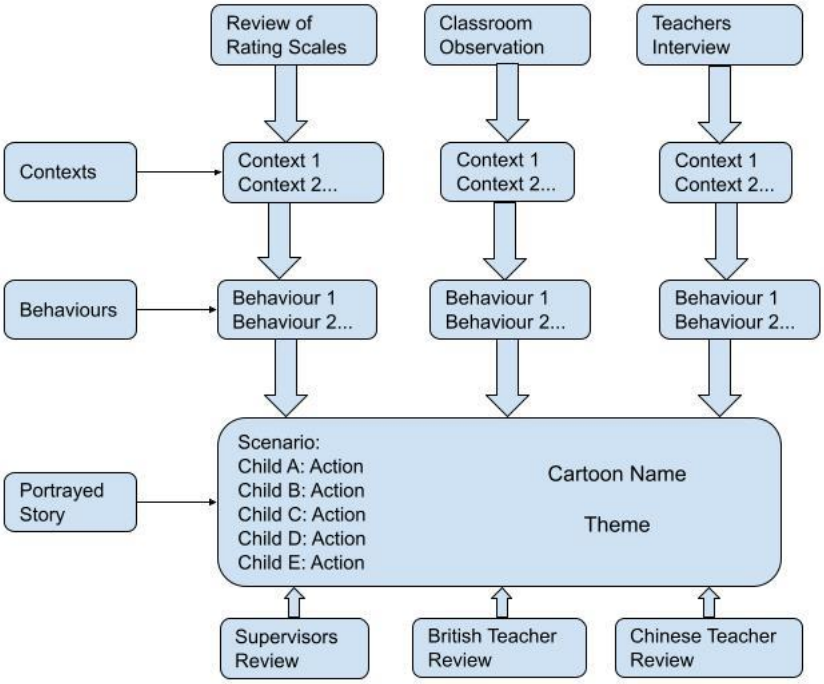
7.2.2 Writing Cartoon Scripts

The second step is Writing Cartoon Scripts. A script can be defined as 'the written text of a film, play, broadcast, or speech' (Cambridge Dictionary). Writing a script is akin to directing a film, where every detail the creator wishes the audience to visualise is conveyed through words. There are two primary methods for writing cartoon scripts: the master scene method and the shot-by-shot method. The master scene method involves setting up the scene using a timeline, with the remaining actions described without specifying camera angles (Scott, 2004). On the other hand, the shot-by-shot method necessitates describing each shot, including camera angles from the audience's perspective (Scott, 2004). Given that I do not possess the expertise of a professional film director or cartoon animator, utilising the shot-by-shot method would require a deep understanding of camera techniques, making it a complex choice for a beginner. In contrast, the master scene method aligns more with my objectives. Through this approach, I can establish the scenes and actions for the cartoons, essentially crafting the narratives and behaviours I intend to depict. However, I've allowed room for professional animators to transform these concepts into the final animations.

The structure of a script should exhibit logical and coherent organisation. The script's structure serves as the foundation upon which the creator can build the call to action. It is essential for the narrative structure of a script to follow a logical pattern, ensuring that it conveys meaning effectively and captivates the audience. Furthermore, this structured format aids animators in translating the script into animations. A standard and effective script structure typically

includes three key elements: location, action, and dialogue. The location is presented at the beginning of each scene. Since the objective is to portray ADHD symptom behaviours within a classroom setting, the location is consistently a classroom, with scenes representing various contexts of ADHD symptoms and actions representing the behaviours. All of these elements have been derived from the findings of three distinct research methods: a comprehensive literature review of rating scales, classroom observations, and interviews with teachers. This process is visually demonstrated in *Figure 3*.

Figure 3 The cartoon scripts writing procedure



The purpose of developing these cartoons is to assess teachers' attitudes towards ADHD behaviours. Therefore, the scripts must accurately depict the contexts and behaviours of ADHD that occur in a classroom setting. These portrayals were based on findings from three different studies: a review of rating scales, classroom observations, and interviews with teachers. These three research angles provided insights akin to different camera angles capturing the behaviours

associated with ADHD symptoms in the classroom. They served as a rich source of material for developing the scripts.

To ensure that the survey remains manageable within a 30-minute timeframe, it's important that the cartoons strike a balance. They should be concise enough for teachers to watch, understand, and rate effectively, yet not so brief that they fail to convey the intended messages clearly.

Furthermore, considering cost-effectiveness in animation, the ideal length for these cartoons falls between 20 and 45 seconds. Animators will use these scripts as the basis for the final animations, so it's imperative that the wording in the scripts is simple and clear. According to Jeff (2004), for a 45-second cartoon, a script should contain roughly 90 to 120 words.

To ensure that the scripts are universally understandable to both British and Chinese teachers, draft scripts were reviewed by supervisors, two Year 2 English teachers, and two Year 1 Chinese teachers. Their feedback and approval confirmed that the developed scripts align with the project's initial concepts. Therefore, scriptwriting adhered to these principles:

- 20 to 45 seconds of an animated cartoon
- Six cartoons (six symptoms)
- Six characters (a teacher, five cartoon children)
- Five actions for each cartoon (four ADHD behaviours, one typical behaviour)
- 90 to 120 words for each script
- A consistent and structured format of scripts (Cartoon Name, Theme, Scenario, Child A, Child B, Child C, Child D, Child E)
- Reviews

7.2.3 Classroom Observation

To animate ADHD behaviours in a classroom setting effectively, it's essential to acquire more context about how these behaviours manifest within that specific environment. However, the behaviours described in rating scales lack the context of their occurrence in a classroom setting. These behaviours are general and encompass a wide range of age groups among children and adolescents. Notably, different age groups exhibit varying levels of ADHD behaviours (Larrson et al., 2004). For instance, younger children tend to be more restless and excitable than their older counterparts (Hurtig et al., 2007).

Furthermore, diverse age groups of children engage in different levels of school activities. When teachers are asked to assess a child's behaviours within their classroom, their interpretation of these behaviours should align with the age group they are teaching. Consequently, their understanding of ADHD behaviours, their acceptance, and the context in which these behaviours occur can significantly impact their judgments. Additionally, different cultures have distinct classroom settings, activities, and teaching methodologies. As a result, the context in which ADHD behaviours manifest may vary across different age groups and cultures. To comprehend the context of these ADHD behaviours in British and Chinese schools, classroom observations from both countries become imperative.

Furthermore, observation constitutes a crucial step in scripting for animation. Observation represents a foundational human social activity and is also a systematic technique for collecting and analysing information about people's behaviour in social research (Cohen et al., 2018). This methodology involves systematically observing and noting people, events, behaviours, settings, artefacts, routines, and more (Simpson & Tuson, 2003). It enables researchers to gather 'live' information from naturally occurring situations and provides firsthand insight into

what transpires in a particular context. Observational research is valuable for capturing non-verbal behaviours in natural settings, as noted by Bailey (1994).

Observation-based research is a well-established practice in educational research, with classrooms being the most common sites for such studies. This method is widely employed to monitor the effectiveness of instructional practices in schools and address inequities. Generally, there are two types of observation: experimental and naturalistic. In experimental observation, researchers set up observation sites and carefully selected participants. Observations are structured into predetermined categories and coded using measurable and countable data, which can then be subjected to statistical analysis. In contrast, naturalistic observation does not control the observed environment; it seeks to observe what occurs 'naturally.'

In my research, naturalistic observation aligns with my objectives since I aim to observe how ADHD-type behaviours naturally manifest in Chinese and English school classrooms. Therefore, I conducted naturalistic observations, personally visiting primary schools in both England and China and directly observing children's behaviours. These observations focused on gathering 'live' information about the contexts in which these ADHD behaviours occur and how they manifest within the classroom setting. I maintained observation diaries containing data on timetables, activities, and observed behaviours.

The selection of the two schools was facilitated by my existing connections. The headteacher of the Chinese school was a former colleague, and I had previously volunteered at the British school. This ensured easy access to my research. I initiated the process by sending consent forms to the participating schools, aiming to minimise any interference or distraction caused by my presence during observations. I explained my research purpose and clarified that my

observation did not pertain to their teaching methods. To mitigate potential interference with the children's behaviour due to my presence, which could lead to altered behaviour, I spent four days at the Chinese school and only recorded observations on the last day when the children had become accustomed to my presence. In the British school, the children were already familiar with me from my prior volunteer work, further reducing potential interference with their behaviour.

The Chinese Primary School Introduction

The Chinese primary school I visited is a relatively small institution affiliated with a university. It exclusively enrols children of university staff members. This primary school operates as a public institution funded jointly by the government and the university. It adheres to the national curriculum, and all students undergo national examinations in Year 6 to determine their choice of high school. The primary school comprises two Year 1 classes, each accommodating 29 students. Throughout indoor lessons, students were typically required to remain at their desks, except during physical education (PE) classes.

The national curriculum encompasses subjects such as Chinese, Mathematics, English, Music, Arts, and PE. The primary mode of instruction predominantly consists of teacher-led lectures, students copying what the teachers demonstrate, and completing exercises with pen and paper. Group activities, gaming, and personalised learning approaches were not observed within the curriculum. *Figure 4* shows the Y1 Chinese classroom setting.

Figure 4 The Chinese classroom setting of primary school

Moreover, children could not access educational applications on computers or iPads to aid their learning. All the children were taught in the same manner and at the same pace. There is no uniform requirement in primary school. The school begins at 8:00 a.m. and ends at 4:00 p.m., with all classes following the same timetable. There are four lessons in the morning and two lessons in the afternoon, each lasting 40 minutes. There is a two-and-a-half-hour lunch break, during which the children go home for lunch and a midday break.

I spent four days in this primary school, sitting at the back of the classroom. On the first day, the teacher introduced me to the children. They were aware that I was conducting educational research in their classroom but were unaware that I would be observing their behaviour. Nonetheless, they behaved very well on the first day. During the lessons, some children would glance at me to see if I was watching them when they did something 'wrong'. If I made eye contact, they would cease their actions. The teachers even mentioned that the children's

behaviour was better than usual due to my presence. I learned from the teachers that the school regularly monitors teachers' teaching and children's behaviour in the classroom by having staff members sit at the back of the classroom and observe. On the second day, the children showed interest in me. Some asked me about my purpose in their classroom, while others shared stories about their new pencil cases, dresses, and school bags. A young boy who sat next to me even crafted two toy cars using rubber and pencils, and he wanted to play with me. Recognising the impact of my presence on the children in the classroom, I only began recording their behaviour on the last days of my stay.

All the observational data was recorded in observation diaries, including timetables, activities, and observed behaviours. *Table 6* shows the timetable of classroom activities of the observed Y 1 class on the recording day.

Table 6 The timetable for classroom activities of the Y 1 class in the Chinese primary school

Time	Activities
8:00 - 8:15	Assembly
8:15 - 8: 35	Morning reading
8:40 - 9:20	Math
9: 30 - 10: 10	Chinese
10:10 - 10: 25	Morning Exercises
10:25 - 11:05	Music
11:15 - 11:55	English
11:55 - 14:30	Lunch break
14:30 - 15:10	Math practice
15:10 - 15:25	Eye exercises

15:25 - 16:05	PE
16:05	off school

Chinese Classroom Observation Diary

Location: Year 1 Class 1 Age 7 student number: 29

Time: 8:00-8:15 Assembly

Activity: The kids assembled in the courtyard of the school. The teacher asked kids to stand in line, stand still, be quiet, and not talk, laugh, or move around. The teacher asked the kids to check whether they had brought all the school kits. Boys were in one line, and girls were in another line. Kids were signed in by answering when their names were called. After the Assembly, kids were asked to enter their classroom in line.

Observed Behaviours:

● Most of the kids could not stand still. They were jumping, copying each other's actions, laughing, walking around, talking, rocking from side to side, squatting and playing with dry leaves, twigs and stones on the ground.

● Girls were playing with their clothes and dancing.

● Boys were tickling, pushing, and hugging each other.

● When the teacher asked to return to the classroom in line, two or three children could not follow the instructions. Instead of following other kids and walking back to the classroom, they were standing there.

Time: 8:15-8:35 Morning reading (Chinese)

Activity: The teacher read an article in Chinese sentence by sentence. Kids were asked to repeat the sentence by following the teacher.

Observed Behaviours:

● Leaving the seat to look at what another kid was doing and talk to him.

● Picking something out from their pockets and playing with it (stickers, a piece of paper, a plastic stick).

● Playing with their textbooks (looking at the pictures in the books, drawing, turning pages, folding pages).

● Playing with everything on their table (pencils, pencil bag, ruler, rubber), playing with a piece of paper (fold it, tear it).

● Looking around, leaning on the table, putting the foot on the chair, and playing with the shoes.

● Moving the table forward and backwards.

Time: 8:40-9:20 Maths

Activity: The teacher was teaching how to count from 1 to 10 and do addition and subtraction by counting forwards and backwards. The teacher demonstrated how to count different animals in one picture, and students were given pictures with different animals. They were asked to count the animals in the picture and write down the number of each animal. The teacher asked the children questions about the numbers they wrote down. The teacher demonstrated how to do addition and subtraction by counting forwards and backwards. The teacher used colourful plastic sticks to illustrate the connection between counting forward and addiction and the connection between counting backwards and subtraction. Then the teacher gave each kid a bundle of plastic sticks to practice and asked the kids the answers to some questions about addition and subtraction.

Observed Behaviours:

● Some kids could not be quiet, even when the teacher asked them to stop talking. They were talking to each other and talking to themselves.

● The teacher asked the kids to wait for their turn to answer the questions and raise their hands after she finished the question. Some kids could not wait and always blurt out answers before the teacher finished the question or before their turns.

● Some kids also could not be quiet when other kids were answering questions; they were talking to other kids, shouting out their answers when they disagreed with the kid who just answered the question.

● Some kids were not listening to the teacher. They were daydreaming by looking out the window or staring at their textbook.

● A boy made a toy car with paper and rubber and showed to other kids and played with it on another kid's table.

● A boy interrupted the teacher and asked for permission to go to the toilet (school rules: go to the toilet during the break time and no permission to go to the toilet while lessons, no drinking and eating).

● Some kids were leaving their seats to drink water, taking snacks out of their pockets, and eating.

● Some were leaving their seat to chat with other kids or walking around the classroom.

● They were playing with their stationery.

● Some kids could not wait for their turns when the teacher handed out the plastic sticks. Some left their seat and walked to the teacher to ask for the sticks. A girl was leaving the seat and grabbing the plastic sticks from another kid's desk. A boy was walking to the teacher and grabbing a bundle of plastic sticks from the teacher's basket.

Time: 9:30-10:10 Chinese

Activity: The teacher was teaching Pinyin on the blackboard. She was demonstrating the pronunciation and handwriting of some Pinyin. She was showing how to join these Pinyin together to pronounce words. The kids were asked to pronounce the words by following the teacher. The teacher asked questions about the pronunciation of new words. The kids were asked to raise their hands and wait for their turn to answer these questions.

Observed Behaviours:

● Doodling on the textbook.

● Being distracted by the PE lesson of another class outside the classroom and looking out of the door.

● Keeping dropping things on the floor and picking them up.

● Laughing at other kids when they did not get the correct answer to the teacher's questions. Moving their chairs forward and backwards.

● looking for books and stationery in their drawers.

● Making silly faces to make other kids laugh.

● Using pencils and rulers as swords to play fight with each other.

● Making funny noises and humming.

● Taking a straw out of the drawer and putting it into the mouth. Spitting saliva into the straw and dropping it to the table and blowing bubbles into the saliva.

Time: Morning Exercises: 10: 10-10:25

Activity: All the kids assembled and lined up in the school courtyard. The kids were asked to stand still and keep arms-length space between each other when they lined up. After they lined up, the broadcast gave instructions and music. The kids were doing exercises by following the PE teacher's demonstration and the instructions from the radio.

Observed Behaviours:

● They stayed in the classroom drinking water and playing with their water bottle. They forgot to go out to the assembly. The teacher came in to remind them.

● They were hugging and pushing each other in the line.

● They were running forwards and backwards all of a sudden.

● They were jumping, twisting body, bowing down and playing with insects, twigs and dry leaves on the ground.

● They were talking and laughing instead of following the PE teacher's demonstration.

● Some kids could not follow the instruction and demonstration of the PE teacher to move their bodies. Their movements were either slower than others or lost completely.

Time: Music: 10:25-11:05

Activity: The kids were sitting in the music room in pairs. The teacher was playing the piano and singing. Every kid has a musical notation book of songs. They were asked to follow the teacher's singing and the notation.

Observed Behaviours:

● Taking off the shoes and playing with them (putting the pencil in the shoes).

● Keeping dropping the book on the floor and picking it up.

● Talking to each other and whispering in the ears.

● Playing with the musical book (doodling on the book, reading other pages, folding the page corner).

● Putting a foot on the chair and playing with the shoelaces.

● Knocking the chair with his back repeatedly.

● Playing with a pencil and a rubber (digging holes in the rubber using a pencil, doodling on the rubber).

● Playing with a pencil sharpener and the

● Talking to each other and laughing.

Time: English 11:15-11:55

Activity: The teacher was teaching some English words and short sentences. The kids were asked to repeat these English words and sentences by following the teacher.

Observed Behaviours:

● Playing with the trousers belt.

● Playing with a plastic straw (spitting saliva into the straw when it was going to drip, sucking it back).

● Cutting the rubber by using a ruler.

● Standing on the chair.

● Playing with the plastic book cover.

● Suspending herself between two tables.

● Repeating these English words and sentences but not paying attention at all.

● A butterfly flew into the classroom. Some kids were watching, shouting out and trying to catch it. A boy was chasing the butterfly out of the classroom.

● Poking the nose with fingers, digging out booger, and playing with it like playdough on the desk.

Time: Maths exercise 14:30 - 15:10

Activity: The kids should follow the teacher's instructions to finish the exercises in their math book.

Observed Behaviours:

● Doing exercises without following the teacher's instructions.

● Resting the head on the desk and daydreaming.

● Reading other books instead of doing Maths exercises.

● Lying on the floor and reading the book.

● Playing stationary in the drawer.

● Drinking.

● Leaving the seat to talk or look at other kids.

● Using a pencil to knock the table or book and making strange noises in the mouth.

● Most kids could not focus on tasks for the whole lesson. Some kids finished the tusks quicker than others. Then they got bored and lost interest in the tasks. Others could not follow the instructions of the teacher, and then they got lost and lost focus.

Eye exercises:

Activity: Eyes exercise requires the kids to close their eyes and exercise by following the instructions on the radio.

Observed Behaviours:

● Doing eye exercises with eyes open.

● Reading a book while doing the exercise.

● Not following the instruction and doing it very quickly or slowly suddenly.

● Playing with a girl's ponytail and using a pencil to poke her back.

Time: PE 15:25-16:05

Activity: The kids were learning how to assemble, line up and do exercises.

Observed Behaviours:

● When the teacher asked the kids to stand in lines, they pushed and dragged each other and moved around.

● Distracted by other kids in the playground.

● Some kids could not follow the teacher's instructions to move their bodies and could not coordinate their body movements. They confused with left and right side movements. They could not stand still when the body moved and moved exaggeratedly. They also gave up quickly and played on their own.

Activity: Break time 10 minutes between each lesson.

Observed Behaviours:

● Interrupting other kids' drawings on the blackboard (wipe or doodle on it).

● Taking away the things the other kid was playing with without asking.

● Chasing, pushing, dragging, shouting and screaming at each other in the classroom.

● Playing with classroom doors. One was pushing to close, and another was pushing to open. Once the door was locked, the one who was locked outside was banging and kicking the door.

The impression gathered from the observation is that inattentive, hyperactive, and impulsive behaviours were common in the classroom. Some children displayed these behaviours more

frequently than others. For most children, these behaviours occurred occasionally and did not hinder their learning. Additionally, it was observed that Chinese teachers primarily focused on advancing their lecturing rather than addressing these behavioural problems. The teachers were often under pressure to make progress in teaching their subjects, resulting in intensive and predominantly lecture-based teaching.

The learning approach leaned towards teacher-centred rather than child-centred learning. All the children were taught in the same manner and at the same pace. The primary form of interaction between teachers and children involved asking and answering questions, with teachers usually posing the questions and children responding. However, it was apparent that lecturing did not suit all children, as many became bored and lost focus during lessons. The most frequently observed behaviours, as documented in the diaries, were inattentive. Nevertheless, teachers seldom invested time in managing these behaviours as long as the children remained quiet and did not disrupt others. Notably, children who sat at the front of the classroom exhibited better behaviour. Conversely, those seated at the back displayed more ADHD-type behaviours compared to their front-row peers. Two children situated at the back of the classroom appeared to struggle with remaining seated and staying focused on their learning. Unfortunately, they received minimal attention from the teachers, primarily being located in the back corner of the classroom.

Primary schools have undergone significant changes since my own attendance 37 years ago. During my time in primary school, teachers were stricter in addressing children's behaviours and dedicated considerable time to discipline. We had to maintain a specific posture at our desks, with any deviation resulting in bamboo stick beatings or scolding. Misbehaviour incurred scolding or physical punishment. However, the teaching style was not as intensive as

it is today. The teachers at the school I visited appeared to be more lenient and positive when dealing with misbehaving children. They typically offered gentle reminders to misbehaving children and encouraged good behaviour rather than resorting to scolding and punishment to curb misbehaviour.

The English School Introduction

The English primary school I had permission to visit was a small church infant school that only had children from Reception to Year 2. The state funded the school. There were only 59 children at the school, and the Year 2 class had 18 kids. There was a teacher and a teaching assistant in the Year 2 class. The school started at 9:00 in the morning and finished at 15:05. All the children were wearing school uniforms. Year Two covered the subjects within the Primary National Curriculum: English, Mathematics, Science, Computing, History, Geography, Music, Art, Design Technology, Physical Education, Religious Education, and Spiritual, Moral, Social, and Cultural Development (S.M.S.C.). The children developed their knowledge, skills, and understanding through various experiences, opportunities, and activities. The teachers operated a thematic approach to the curriculum through whole-school cross-curricular topics. The emphasis was on active learning and providing memorable experiences for the children. The learning activities included group work, gaming, online applications for assisted learning, outdoor exploration, crafting, and art design. There was a large playgroun*d* with a timber trail, football pitch, seating area, wildlife garden, pond, and aviary for hens. *Figure 5* shows Y2 classroom setting of English primary school.

Figure 5 Y2 classroom setting of English primary school

All the observational data was kept in observation diaries, including timetable, activities and observed behaviours. *Table 7* shows the timetable of classroom activities of the observed Y 2 class on the recording day.

Table 7 The timetable for classroom activities of the Y 2 class in the English primary school

Time	Activities
8:45 - 9:00	Assembly
9:00 - 9:10	Sign in
9:10 - 10:00	Math
10:00 - 10:15	Church Assembly
10:15 - 10:30	Play time
10:30 - 11:00	Phonic
11:00 - 12:00	Literacy

12:00 - 13:00	Lunch break
13:00 - 14:00	Science
14:00 - 15:00	Visiting the allotment

English Classroom Observation Diary

Location: Year 2 Age 7 student number: 18

Time: Assembly 8:45 - 9:00

Activity: The kids arrived at school and lined up at the playground.

Observed Behaviours:

● Girls were playing with their skirts and socks. Playing with each other's hair.

● Boys were playing with twigs and small stones on the ground.

● Tickling each other.

● Running into other class's line.

● Playing with another kid's PE bag

● Talking and laughing with each other

● Throwing book folders to each other

● When they walked to the classroom, some kids were pushing each other. A boy was looking at the kid behind him, and he walked into the kid in front of him.

Time: Sign in 9:00 - 9:10

Activity: All the kids were sitting on the floor facing the teacher. The teacher called the kids' names one by one. The kids were asked to raise their hands and say 'Yes' when their names were called.

Observed Behaviours:

● Was not paying attention and did not hear the teacher calling his name. He was reminded by the kid who was sitting next to him. He answered 'Yes' but did not raise his hand.

● A kid was late for school. She looked a bit upset and tired and came into the classroom with the teaching assistant. It took her a while to settle down. She did not bring her school folder to school.

Time: Maths: 9:10 - 10:00

Activity: The kids were sitting on the floor and listening to the teacher explain how to. The teacher asked the kids questions. Then the kids were separated into groups for different tasks. A group of kids were using Maths Apps on the computer for exercises. A group of children were counting and drawing. A group of children were given different shapes of plastic toys to group them and count the number. A group of children were working with the teaching assistant.

Observed Behaviours:

- Playing with shoelaces.
- Playing with a girl's hair who was sitting in front of him.
- Not sitting appropriately. Stretching the legs and lying on the floor.
- Moving forwards and backwards while sitting on the floor.
- Fidgeting with a small piece of playdough, squeezing and rolling it in the hands.
- Daydreaming, not listening, empty-looking
- Playing games instead of doing Maths exercises on the computer.
- Losing interest quickly and playing with a mark pen doodling when working with another kid.
- Being distracted by what the kids were doing at the next table.
- Walking around the classroom and playing with stickers on a book.
- Having difficulty understanding the teaching assistant's instruction. After the teaching assistant left the table, instead of doing what the teaching assistant had explained and showing him what to do, the kid looked confused and lost interest quickly.

Time: Church Assembly 10:00 - 10: 15

Activity: A priest came to the school for the Church Assembly. All the kids were gathered in a room. The priest has a doll called Mr Teddy. The priest and Mr Teddy were talking about the story of Jesus. The story was about what happened when Jesus was born. Mr Teddy asked the kids questions and interacted with the kids. Then all the kids read a poem about Jesus' birth by following Mr Teddy.

Observed Behaviours:

● Not listening, not interacting with Mr Teddy when other kids were interacting and laughing.

● Tickling the kid who sat next to him

● Being distracted by the lights from a projector.

● Giggling and laughing with other kids

● Playing with the plants in the room.

● When the priest asked who wanted to take Mr Teddy home tonight, instead of raising a hand and waiting for his turn, a kid who sat at the front stood up and grabbed Mr Teddy in his arms. Some kids were shouting, and some who sat at the back raised their hands and stood up.

Time: Playtime 10:15 - 10:30

Activity: All the kids were outdoors and playing with the facilities in the playground. There were slides, a long climbing wall, a wooden shelter, a wooden bridge and a sandbox at the playground.

Observed Behaviours:

● Excessive running like a little rocket without stopping. Knocking other kids by arms when he was running.

● Could not wait for turns when playing. When other kids were waiting for their turn to slide down from the slide, he walked up to the slide and pushed other kids away.

● Trying to climb up the shelter.

● Not playing with other kids. Playing with the facilities but losing interest quickly.

● Destroying the sandcastle which other kids built.

Time: Phonic 10:30 - 11:00

Activity: The teacher reviewed the phonics they have learned before with the kids. The teacher held phonic cards and asked the kids to pronounce the phonic. The teacher also asked the kids to find out the words pronounced phonic. The kids were sitting on the floor. They were required to raise their hand and wait for their turn before answering the questions.

Observed Behaviours:

● Disturbing the teacher's talk with an unrelated topic.

● Burst out the answer without raising hand and wait for the turn.

● Not listening and playing with a sticker on the carpet.

● Whispering to the kid who sat next to her and laughing.

Time: Literacy 11:00 - 12:00

Activity: The lesson is about how to write an adventure story. All the kids were sitting on the floor. The teacher used a display poster to explain the definition of an adventure story and the key features of an adventure story. Then the teacher asked the kids to give some keywords and settings of an adventure story. Then the kids worked in groups by sitting at the table to write an adventure story and draw the characters and settings by sitting. There was a volunteer came in to help the kids with their reading. The kids went one by one to volunteer to read a book with the volunteer at the classroom reading corner. The kids would choose a book they like in their reading level and read it to the volunteer.

Observed Behaviours:

● Playing with the hair in the mouth. Winding the hair on the fingers.

● Cannot sit appropriately. Putting a foot on the chair, fidgeting on the chair by moving the body forwards and backwards.

● Playing with a little box by spinning it on the table.

● Having difficulty reading the book.

● Having difficulty of focus on reading, wanting to talk about something irrelevant to the reading.

● Keep looking for pens, glue, scissors and so on while working on the tusks.

● Could not get the work finished. He needs the teacher and teaching assistant to remind him what to do next.

● Winding around the classroom and doodling on the blackboard.

● Could not work with other kids in the same group, looking confused.

● Swinging on the chair

Time: Science 13:00 - 14:00

Activity: All the kids were sitting on the floor. The topic was plants and living things. The school has some chickens that have laid eggs, which were hatched in the school. The chicks were kept in the village allotment. The teacher showed the kids pictures of hens and eggs, how the eggs were hatched and the chicken cage in the village allotment. The teacher and kids discussed what chickens eat and what vegetables and plants the village allotment has. A fly was in the classroom. The fly was flying against the window and trying to fly out.

Observed Behaviours:

● Making a funny face to entertain the kid sitting next to her

● Leaning the body side to side while sitting on the floor.

● Talking to the kid who sat next to him

● Daydreaming and not interacting with the teacher.

● Shouting out the fly

● Distracted by the fly's noise and watching the fly all the time.

● Trying to catch the fly with hands.

Time: Visiting the village allotment 14:00 - 15:00

Activity: Kids held each other's hands in pairs and walked the village allotment with the teacher, teaching assistant and a school volunteer. The kids visited the hen house where the school kept all the hatched chickens there. The kids collected the eggs in the hen house. The teacher and teaching assistant showed the kids all kinds of vegetables, fruits and flowers people grow in the allotment.

Observed Behaviours:

● Not listening and falling over even the teacher has reminded.

● Always fell behind the group.

● Not paying attention to what the teacher said. Distracted by other things on the way to the allotment.

● Getting upset quickly (complaining about another kid not doing what he wants him to do)

● Running excessively on the pedestrian

● Climbing up the flower beds.

It was a very interesting experience to observe children's behaviour in an infant church school in England. My first impression was that most of the children enjoyed their learning because the children had the autonomy to learn. It was child-centred learning rather than teacher-centred learning. Teachers only guided and supported their learning by assigning different learning activities to the children, which suited their learning abilities. The learning involved intensive interaction between teachers and children, as well as among the children themselves. There was no lecturing but mainly discussions between the teacher and the children. Children also asked questions, which I did not observe in the Chinese classroom.

My second impression was that the learning activities were diverse in the classroom. Even when learning about the same topic, children were engaged in different activities. Some were working on tasks with other children, some were using educational apps to learn, and some were interacting with the teacher or the teaching assistant. Children were constantly switching between different activities to prevent boredom.

My third impression was that teachers not only focused on children's learning but also focused on encouraging good behaviour. This church's infant school prioritises good manners and behaviour in children. Teachers were attentive to misbehaviours and were willing to manage these behaviours in the classroom. Furthermore, teachers appeared competent in handling these misbehaviours in flexible ways, such as offering reminders, providing extra support, utilising learning apps, and making use of school facilities. The teacher had a reward system in place to encourage good behaviour, involving stickers, the opportunity to care for school pets (stick insects), and taking Mr Teddy (a soft toy) home. Given the small class size and the presence of both a teacher and a teaching assistant, the children's misbehaviours were well managed.

7.2.4 Teachers' Interview

Although through observation, I have vivid diaries of what happened in the classroom, the researcher observes and interprets what happens in the observation method. Therefore, there is always the potential for observer bias and subjectivity. While my observation followed the diagnostic criteria for this research, I have no experience in teaching children who are 7 years old. I do not know the teachers' opinions about these behaviours. I was sitting in the classroom to observe children's behaviours while the teachers were under the heavy duty of teaching and managing these behaviours in the classroom. The observer and the teachers have different roles in the classroom. The observer could assume something is 'wrong' because the children being observed are doing things differently from what the observer is used to, expected, or accepted.

However, teachers may have completely different opinions. Moreover, the researcher's gender, race, and social class could potentially influence the observation and interpretation of the observed behaviours. Therefore, it is essential to have another method to compensate for potential bias and determine teachers' opinions about ADHD behaviours in their classrooms.

The interview method provides an opportunity to explore participants' experiences and understandings. It involves a purposeful interaction between researchers and participants, enabling the researcher to learn what the participants know about a topic, discover and record their experiences, thoughts, feelings, and the significance or meaning they attach to them (Arthur et al., 2017). The interview grants access to a person's thoughts, allowing for the collection of information about what a person knows (knowledge or information), what a person likes or dislikes (values and preferences), and what a person thinks (attitudes and beliefs) (Tuckman, 1972). This method may also be used in conjunction with other research methods to follow up on unexpected results, validate other methods, or delve deeper into respondents' motivations and reasons for their responses (Kerlinger, 1970). Interviews can help test hypotheses, suggest new ones, or serve as an explanatory device to help identify variables and relationships (Cohen et al., 2011).

As the DSM-V criteria have been criticised for describing symptoms that are typical behaviours not exclusive to children with ADHD but common in all children to some degree, and for lacking context, it becomes challenging to identify a child with ADHD based solely on these behaviours. Moreover, judgments based on different interpretations in various contexts can be subjective. To understand how ADHD behaviours manifest in the classroom and in what contexts teachers perceive these behaviours, conducting interviews to gather teachers' opinions about these ADHD symptoms is an essential method.

Additionally, this method can validate the behaviours observed in classrooms. The purpose of the interview was to determine which behaviours teachers consider to be ADHD behaviours and in what context they believe these behaviours occur. The interviews followed a semi-structured format, with structured questions aligning with the 18 ADHD symptoms outlined in the DSM-V. Unstructured questions emerge during the conversation to seek clarification and further insights. While the interviews were recorded, they were not transcribed. Instead, information about the behaviours and contexts was extracted and categorised under the 18 core symptoms of ADHD through repeated review of the recorded material. Two Year 2 teachers were interviewed, one from China and the other from England.

Examples of questions:

Could you please give me some examples of behaviours of children that often has trouble holding attention on tasks or play activities in the classroom?

关于儿童在教室里无法集中注意力学习的行为，请您能举一些具体的例子吗？

Could you please give me some examples of children's behaviours that often interrupt or intrude on others in the classroom?

关于儿童在教室里干扰其他同学学习或者玩耍的行为，请您能举一些具体的例子吗？

Chinese Teacher's Interview

The teacher from China had eight years of teaching experience with Year 2 children. She taught the subjects of Chinese and PE. She had heard about ADHD. She mentioned that she had taught children she was concerned may have ADHD. However, she did not know whether the children had been diagnosed because teachers could not refer children to be diagnosed. In China, there was no educational reinforcement of ADHD diagnosis. It depended on the parents' decision. Moreover, the parents might not have been transparent with teachers and schools about the diagnostic results. Meanwhile, teachers did not have external support for their teaching if they had children with ADHD in their classrooms. The study extracted children's behaviours from

the teacher's interview into a categorised table. *Table 8* shows all the behaviours extracted from this Chinese teacher's interview.

Table 8 Children's behaviours extracted from the Chinese teacher's interview

Inattention	**Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.** • The handwriting of the Chinese characters is in a mess. Some parts of the characters they write about are often missing • When they learn Pinyin, they always get some pronunciations mixed up, e.g. lie and lei, an and ang, en and eng. • They make the same mistakes repeatedly, and even you have reminded them many times. The teacher asked them to rewrite the characters which they had written wrong. They wrote them correctly at the beginning and then wrote them wrong again. **Often has trouble holding attention on tasks or play activities.** ● The lessons are 40 minutes. The child only focuses for 5 or 10 minutes at the beginning of the lesson. Then they start looking around, daydreaming, singing, self-talking, reading other books, playing with stickers, and clothes tags, writing, and drawing. **Often does not seem to listen when spoken to directly.** ● When I talk to them, they do not look at me, or they look at me for a few seconds, and then they lose eye contact with me, and I know they are not listening. They look around or their head down and look at the desk or play something in their hands but do not listen to me. ● They do not show facial expressions of interest when I talk to them. Sometimes, the kids are not even upset when they are told off. I feel that they are not in the same world as me. ● When I ask them questions, I do not get answers from them. I do not know whether they cannot hear me, they do not understand my questions, or they do not know the answers. It is hard to get them to interact with me.

Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).

- When the teacher gives instructions such as 'please open the book, turn to page 27, look at the questions and write down the answers after the questions', they cannot follow the instructions. They will open the book and then sit there and look around. They do not know what to do next. I have to walk to them and repeat my instructions or turn the page and point to them the questions.
- Sometimes, they cannot finish their homework if it is a little bit complicated for them. For example, if the homework is writing characters, they can finish it. If it is a topic writing about their holiday, they either can't finish it, or the quality is very low. They need lots of supervision and help from their parents to finish their homework. The parents report that it is a nightmare to help them with their homework.
- They can't follow the instructions of what they need to do with their homework. I have to send the homework requirement to their parents to help the children with their homework.

Often has trouble organizing tasks and activities.

- Children in pairs have classroom duties in turn, such as cleaning the blackboard and floor, helping teachers carry teaching staff, collecting homework, tidying up the classroom after school and so on. The teacher can’t put them with other children to do the duties because they don’t know what to do. They might manage to do one or two tasks but can’t work with other children to do all the jobs. Most of the time, it ends up with the other child doing most of the tasks.
- They usually play on their own at playtime.
- It is difficult for them to organize their schoolbooks and stationery for the lesson. They forget to prepare things for the next lesson, and Their desks are in a mess.

Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).

- When they feel that the tasks are difficult, they will just sit there doing nothing or lose interest and play something else.
- When they have the lessons they are interested in, they can focus for longer. However, if they are not interested in the subject, they are not focused and do not interact with me. They don't do what you want them to do.

Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).

- They forget to bring their books, homework, craft kit, PE kit to school
- They lose their pencils, books, pencil box and they have to borrow from other children they need it.
- When their parents come to collect them from school, they forget to take their things (eg. books, homework sheet, pencil box) to home.

Is often easily distracted

- They are easily distracted by other children. If other children make noise, they will be distracted to look at them or laugh at them.
- They are easily distracted by the pictures in the book rather than the content. They will look at the pictures, doodle on the pictures rather than follow the teacher's instructions.
- They are easily distracted by what happens outside the classroom (PE lessons of other children, somebody walk pass their classroom)
- They are easily distracted by the bees, butterflies, flies or other bugs flying into the classroom. They will watch them all the time or try to catch, play and even chase them.

Is often forgetful in daily activities.

- They forget daily activities such as eye exercise, morning exercises. For eye exercise, kids are required to maintain in their seats after the

	lesson. Some kids forget this daily activity and go out of the classroom after the lesson. The morning exercises need children to go out to the playground to assembly after the lesson. Some children forget this daily activity, and they stay in the classroom. ● They forget to get everything ready for the next lesson. ● They are late for school. The parents complain that it is very hard to get them ready for school in the morning.
Hyperactivity-Impulsivity	**Often fidgets with or taps hands or feet, or squirms in seat.** ● They cannot sit still at the seats. They are leaning on the table, lying on the chair, putting feet on the chair, kneel on the chair, moving table or chair and making noise, keeping dropping things on the floor and picking up, tapping feet, tapping the table, always playing something in their hands. ● It is hard for them to settle down and walk on a balance beam or pipe. **Often leaves seat in situations when remaining seated is expected.** ● They leave the seat to take or borrow things from other kids or look at what other kids are doing or talk to other kids, or drink, go to the toilet without permission, run out of the classroom without permission, leave the seat to throw rubbish. ● When they get excited or bored, they leave the seat to play, move around in the classroom. **Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).** ● They run excessively at the PE lesson. It is hard to get them to settle down to follow the instructions. ● They climb up the flower raised beds, school gate, fences, basketball stand, chairs, tables, staircase handrails. **Often unable to play or take part in leisure activities quietly.**

- Some kids make noise in their mouth. They talk to other children or themselves. Sometimes, they make a loud noise when they are upset, such as knocking the table with a pencil box.
- They shout when they feel that other children don't play with them, or don't do what they want the other children to do.
- They shout and scream a lot at playtime.
- They usually end up playing on their own.

Is often "on the go" acting as if "driven by a motor".

- It is hard for them to settle down after the short break between lessons. They still look at the playground or windows. They cannot sit still and keep moving on the chair.
- They play with EVERYTHING (pencil, ruler, rubber, books, pencil bag, pencil sharpener, clothes, shoes, clothes zip, tag, string and belt, a little piece of paper (fold, tear, draw, cut), their saliva and booger.
- If they get told off, they can stop but only for a short period. Then they will start again.

Often talks excessively.

- They like to talk to other kids.
- They talk to themselves.
- When they are interested in the topic, they talk excessively, and it is hard to get them to stop.
- Sometimes they like to talk about something irrelevant to the lesson.

Often blurts out an answer before a question has been completed.

- When they know the answers, they can't raise their hand and wait for the teacher to call them. Instead, they shout out the answers.
- Sometimes, they laugh at other children's answers.
- When they don't get called to answer the questions, they get very upset or angry.

Often has trouble waiting his/her turn.

- If they want something, they have to get it (sweets, practice tools, working sheet etc.). Otherwise, they get very upset and angry. They

will shout, make a big noise, such as banging the book on the desk, kicking the desk.

Often interrupts or intrudes on others (e.g., butts into conversations or games)

- They interrupt other kids when they are concentrating on working.
- They took pencils, rubbers or other stationery away from other kids without asking.
- They like to disturb other kids' game or play by rushing into their play, pushing them, throwing things to other kids (a small piece of paper or rubber), tickling them or taking their toys away. It is more like trying to get attention from other children by disturbing or annoying them. They don't know how to invite other kids to play their game and don't know how to join other kids' game or play.
- They like to annoy other kids to chase them by taking away other kids' toys, books or pencil box.
- They disturb the teacher's teaching by asking question irrelevant or playing sickness (stomach, headache).

The Chinese teacher I interviewed is very experienced. She could provide many details about children's behaviours. Although she had heard about ADHD, she was not knowledgeable about all the symptoms and the diagnosis of ADHD. During the interview, she also expressed her struggles with teaching children with mental illnesses or disorders. She had encountered a few children with learning difficulties and disorders during her teaching career. She knew that they were different from other children. However, none of them had a medical diagnosis, nor was she informed. She did not know what type of disorder these children had. Since ADHD is the most well-known children's disorder, she got confused about the symptoms of Tourette's syndrome and Dyslexia with ADHD. During the interview, she described the behaviours of a group of children she thought might have ADHD.

In China, there is a stigma against mental illness or disorders. People with mental illnesses or disorders are perceived as 'crazy,' 'incapable,' or 'unfit' (Yu et al., 2018). The stigma excludes them from professional and social opportunities. Therefore, people with mental illnesses will hide their situation from others. Parents of children with mental illnesses or disorders usually do not want the teachers or schools to know to protect their children from the stigma (Xu et al., 2017). Meanwhile, due to the lack of training and support in teaching children with disorders, teachers do not know how to teach and help these children (Liu et al., 2016). The teacher I interviewed expressed that she got stressed out if she had a child with ADHD in her class for two reasons. One reason was that she thought the child was a bad influence on other children. Another reason was that she perceived herself as failing to discipline the child's behaviour. She believed children's ADHD is mainly due to genetics and parenting style.

English Teacher's Interview

The teacher from England had ten years of teaching experience with Year 2 children. She had taught children with ADHD. She received external support for her teaching because she had a child with learning difficulties in her class. She had also received training in teaching children with ADHD. The extracted children's ADHD behaviours are shown in *Table 9***.**

Table 9 Children's ADHD behaviours extracted from the English teacher's interview

Inattention	**Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.** ● They make careless mistakes and forgets things quickly even though he has learned and been reminded many times (e.g., full stops and capital letters for sentences). ● Often has trouble holding attention on tasks or play activities. ● It is hard to keep their focus on activities. ● Daydream and wonders. ● Requiring the teacher lots of attention to keep him on track.

- Not listening and without eye contact, looking at something else in the room.
- They very short attention.
- They have short memory of remembering information.

Often does not seem to listen when spoken to directly.

- I have to keep him close to me. He sits close to me to make sure he is listening and paying attention to what I am saying.
- I also ask him to look at me in the eyes to make sure I have his attention and listen to me.

Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).

- I have to keep asking him questions to make sure he understands what I am talking about and follows the instructions of the tasks.

Often has trouble organizing tasks and activities.

- He can't work with other children. If he works with another child, he will not do anything. He will let another child do everything. He sits back and watches what other children were doing. He does not have the ability to share his ideas, opinions and feelings.
- He is not a strong character. He lets other children tell him what to do rather than invites other children to play his game.
- He can only manage one simple task which can be finished in a very short period (e.g., 10 minutes). For example, he cannot understand instructions like ’Go to the toilet, wash your hands and sit on the carpet’. I have to break the task into three tasks and give him one at a time. He cannot understand instructions or school tasks because they are too complicated for him, and he cannot hold all the information.
- He has the difficulty of planning, prioritizing things and finishing it (e.g., designing and making a leaflet). He gets frustrated easily by doing things.

	● He has the difficulty of solving problems. **Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).** ● He has no confidence to do things he does not know what to do. ● He needs lots of help from the parents to finish the homework. **Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).** ● He has the difficulty of getting all the equipment for activities, and I have to get everything ready for him. If any material is missing, he will just sit there and do nothing. **Is often easily distracted.** ● He is easily distracted (e.g., building work outside the classroom, diggers, music, new toys, any noise). **Is often forgetful in daily activities.** ● He can remember the daily routine, but it is difficult for him to adapt to any change (e.g., supply teacher, different computers, different seat, fire practice). He gets angry, worried, and panic about these changes.
Hyperactivity-Impulsivity	**Often fidgets with or taps hands or feet, or squirms in seat.** ● Sitting but fiddling with shoes, squirming, tapping on the chair leg ● He dances foot to foot or wiggling when all the children are in line. **Often leaves seat in situations when remaining seated is expected.** ● He cannot sit on his seat for a long time, and he will leave his seat and follow the teacher or wander around in the classroom. ● On the school bus, I have to sit with him to make sure he does not leave his seat. **Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).**

- I have to hold his hands all the way around at school trips. If he sees anything interesting, he will run for it or have a go with it or climb up.

Often unable to play or take part in leisure activities quietly.

- He makes clicking noises with his mouth when he is working and grunts when he is not happy. He makes more noise than other children when he is excited.

Is often "on the go" acting as if "driven by a motor".

- He is very energetic. He gets excited easily and it is hard to stop him.

Often talks excessively.

- If he is interested in the topic, he talks excessively, and it is hard to get him stops.

Often blurts out an answer before a question has been completed.

- He shouts out answers when he thinks he knows the answer.

Often has trouble waiting his/her turn.

- He does not have the ability to do something in turn. When he feels he does not get a go, he would hit another child or say something not nice to another child.
- If he wants something, he cannot wait or ask (toys, computers, swing, ball). He will go and get it.

Often interrupts or intrudes on others (e.g., butts into conversations or games)

- Disturbing other children by chatting with them, throwing rubber to them, tapping another children's chair.
- When he is frustrated, he shouts. When he gets angry, he throws things and swears to me.
- He does not understand personal space. He suddenly sits very close to me and put his hand on my lap.

This teacher has extensive experience in teaching children with ADHD. She provided detailed insights into their behaviours and the context in which they occur. Thanks to her professional

training and support, she feels confident in teaching and managing children with ADHD in her classroom. Importantly, she does not hold negative perceptions of children with ADHD; she views them as simply different. During the interview, she described the behaviours of a boy with ADHD in her class. The teacher observed that, despite facing various challenges, he demonstrated the ability to concentrate and focus when engaged in computer-based tasks. Additionally, he displayed a strong fascination with computing. She believes that, with the right support, this boy has the potential to excel.

7.2.5 The Review of Rating Scales of Children's ADHD

Since there are no biological markers for identifying a child with ADHD, the diagnostic criteria of this disorder are based on behavioural symptoms. The objectives of the assessment of ADHD are identifying the presence of symptoms of ADHD and ruling out other possible disorders, developing future intervention plans, and identifying other possible comorbidities (Barkley, 1998). To achieve these objectives, behavioural rating scales are the most common assessment instruments, which are designed into behavioural checklists to assess a wide range of behaviours and symptoms of ADHD (Demaray et al., 2003).

Although a thorough assessment includes rating scales and other assessment tools such as interviews, psychological tests, educational testing and medical evaluation, behavioural rating scales are essential tools for schools and communities. Compared to less structured interviews, rating scales are better standardised and easily administered, and teachers and parents can complete them. Therefore, they are a more reliable and cost-efficient tool for gathering information from multiple informants across different settings. Due to the advantages of rating scales, they can address the presence of ADHD symptoms and the severity, impairments, and comorbidities of this disorder. Therefore, rating scales are essential and widely used in schools and communities.

Although rating scales have many benefits, they have some limitations. Rating scales rely on the assumption that the informant is familiar with the child's behaviour and understands the questions in rating scales. However, this is not the case all the time in reality. Furthermore, the informant's psychopathology can affect the parent and teacher's perceptions of child behaviour (Anastopoulos & Shelton, 2001). Therefore, subjectivity and rater bias are potential limitations of using rating scales.

Although the DSM provides descriptions of symptoms, behaviours related to each symptom can vary between contexts. Therefore, various diagnostic instruments have been developed to gather information from informants in line with the DSM-V diagnostic criteria for ADHD. In the diagnostic process, teachers, acting as informants, complete a questionnaire that includes a behavioural rating scale related to the observed behaviours of the child under investigation in a school setting. To gain a better understanding of the specific ADHD-related behaviours that teachers are instructed to observe in these rating scales, I investigated several widely used rating scales. This inquiry primarily focuses on teacher rating scales and the items (behaviours) within these instruments. Essentially, this review aims to present a comprehensive overview of how ADHD symptoms are defined and observed by teachers. The examination of these rating scales contributes to ensuring that cartoon scripts accurately portray ADHD behaviours and realistically characterise individuals with ADHD during school screening. Additionally, this review has assisted the researcher in developing a teacher rating scale for assessing children's ADHD behaviours.

There are two types of rating scales: the broad-band rating scales and the narrow-band rating scales. Broad-band rating scales include the DSM diagnostic criteria and a wide range of

behavioural and emotional problems related to ADHD symptoms. Broad-band rating scales have been recommended for screening individuals for ADHD (Barkley, 1998). The narrow-band rating scales, on the other hand, include only the DSM diagnostic criteria and have been recommended for characterising ADHD symptoms in affected individuals (Barkley, 1998). This investigation includes the widely used and researched instruments of both types of rating scales. The review focuses on the instruments shown in Table 10, which are designed for teachers to evaluate children aged 6 years.

Table 10 Reviewed rating scales

Test Name	Items	Age Level	Subscales	Response Format
Conners Rating Scales (CRS-Rs)	28	3-17	Oppositional, Cognitive problems, Inattentive, Hyperactivity	4-point scale based on frequency
Strengths and Difficulties Questionnaires (SDQ)	25	3-16	emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, prosocial behaviour	3-point Likert scale
ADHD Rating Scale-IV (ADHD-IV)	18	4-20	Inattention, Hyperactivity-Impulsivity	4-point scale based on frequency
Attention Deficit Disorder Evaluation Scales (ADDES)	60	4-19	Inattention, Hyperactivity-Impulsivity	5-point scale based on frequency
Vanderbilt ADHD Diagnostic Teacher Rating Scale (VADTRS)	43	6-12	Inattentive and Hyperactive-Impulsive,Oppositional-Defiant/Conduct, Anxiety and Depression	5-point scale based on frequency
Copeland Symptom Checklist for Attention Deficit Disorders	67	Child & Adolescent	Inattention/Distractibility, Impulsivity, Activity Level Problems,	4-point scale based on frequency

			Noncompliance, Attention-Getting Behavior, Immaturity, Poor Achievement/Cognitive & Visual-Motor Problems, Emotional Difficulties, Poor Peer Relations	
Swanson, Nolan, and Pelham-IV Questionnaire (SNAP-IV)	26	5-11	Inattention, Hyperactivity-Impulsivity, ODD	4-point scale based on frequency

Conners Rating Scales-Revised (CRS-Rs)

The Conners' Rating Scales were developed by C. Keith Conners (Conners, 1997). They are among the most popular and well-researched instruments. The questionnaires are designed to be completed by parents and teachers to collect information for evaluating childhood and adolescent ADHD symptoms. The questionnaires can also assess many conditions that coexist with ADHD, such as problem behaviours and comorbid disorders. The questionnaires are easily administered and scored by non-professionals. The CRS-R scales have become the standard for attention and behaviour assessment in children and adolescents. They are widely used by school psychologists, child psychologists, paediatricians, clinics, child protective service agencies, juvenile detention facilities, residential treatment centres, and private practitioners. There are both long and short versions of the rating scale.

The long version for parents has 80 items, and the long version for teachers has 59 items. The short version for parents has 27 items, and the short version for teachers has 28 items. The short version contains subscales for Oppositional Defiant Disorder, Cognitive Problems, Hyperactivity, and other symptoms of ADHD. The items have been adapted to align with the

DSM-IV criteria for ADHD. It provides a fast and reliable screening tool that can help identify children who need additional evaluation and monitor treatment effects on children. Respondents are asked to rate behaviours that have been problematic over the preceding month using a four-point scale labelled with levels of appropriateness (0=Never; 1=Occasionally; 2=Often, quite a bit; 3=Very often, Very frequent). The items of the teacher rating scale are listed in the table of Appendix 5.

Strengths and Difficulties Questionnaires (SDQ)

The Strengths and Difficulties Questionnaire was developed by Goodman (Goodman et al., 1998). It is one of the most widely and internationally used instruments and has been translated into more than 80 languages (Goodman, 2001). The SDQ is designed to be completed by parents and teachers to screen children and adolescents (aged 3 to 16) for emotional and behavioural problems, focusing on their strengths and difficulties in these areas. The questionnaire includes five subscales: the emotional symptoms subscale, conduct problems subscale, hyperactivity/inattention subscale, peer relationship problems subscale, and prosocial behaviour subscale. Three versions of the SDQ are available: a short form, a long form with an impact supplement, and a follow-up form. Each version includes an impact supplement to investigate the length and severity of the child's problems. The SDQ can serve various purposes, including clinical assessment, outcome evaluation, research, and screening. The items of the teacher rating scale can be found in Appendix 6.

ADHD Rating Scale-IV (ADHD-IV)

This rating scale was developed to measure behaviours associated with ADHD and 'to provide clinicians with a method to obtain parent and teacher ratings regarding the frequency of each of the symptoms of ADHD based on DSM-IV criteria' (DuPaul et al., 1998, p. 2). It is a norm-referenced rating scale developed to match the DSM-IV diagnostic criteria of ADHD, including the primarily inattentive subtype, primarily hyperactive-impulsive subtype, and combined

subtype. Consequently, the rating scale consists of 18 items for both the home and school versions, designed for children and adolescents aged 5 to 18. Similar to the diagnostic factors of ADHD in DSM-IV (APA, 1994), the rating scale includes two subscales, Inattention and Hyperactivity-Impulsivity, each comprising 9 items. Teachers can complete the questionnaire in just 5 minutes, reporting the frequency of symptoms observed over the past 6 months using a 4-point scale (Never or Rarely=0, Sometimes=1, Often=2, Very often=3).

The published manual for this rating scale has demonstrated that children's scores on this scale exhibit an adequate to excellent relationship with scores from other similar rating scales, such as Conners' Teacher Rating Scale, Conners' Parent Rating Scale, and ADHD Behaviour Code (Barkley, 1990). Furthermore, this rating scale is sensitive in distinguishing between children with ADHD and those without ADHD, as well as the various subtypes of ADHD. It is widely employed for clinical screening, diagnosis, treatment evaluation, and research purposes. The items of the teacher rating scale can be found in Appendix 7.

Attention Deficit Disorder Evaluation Scales (ADDES)

The ADDES scales are norm-referenced rating scales used to assess the presence of ADHD in children aged 4 to 18 (McCarney, 1995). This tool comprises a home version (ADDES-HV) and a school version (ADDES-SV). ADDES-SV consists of 60 items, which are divided into two subscales: inattentive (29 items) and hyperactive and impulsive (31 items). Each item is scored on a 5-point scale (0 = Not developmentally appropriate for age, 1 = Not observed, 2 = One to several times per month, 3 = One to several times per week, 4 = One to several times per hour). The scale was designed to assess three types of ADHD (ADHD-Inattentive, ADHD-Hyperactive-Impulsive, ADHD-Combined) according to DSM-IV ADHD criteria. Completing the questionnaire typically takes around 20 minutes for a teacher familiar with the target child.

The primary purposes of the ADDES are to: '(a) Screen for characteristics of ADHD, (b) Provide a measure of ADHD, (c) Contribute to the diagnosis of ADHD, (d) aid in the development of intervention goals and objectives for home and school, and (e) Identify intervention activities for ADHD in educational environments' (McCarney, 1995, p. 5). The psychometric properties of the ADDES scales are generally strong. One of the most distinctive features of the ADDES is its connection to intervention. The items of the school version can be found in Appendix 8.

Vanderbilt ADHD Diagnostic Teacher Rating Scale (VARS)

The Vanderbilt ADHD Rating Scales (VARS) were developed to align with the DSM-IV criteria and gather information from parents and teachers to identify children aged 6 to 12 years who may be at risk of ADHD and other comorbidities (Wolraich et al., 1998). These scales come in two versions: a parent report and a teacher report. The Vanderbilt ADHD Diagnostic Teacher Rating Scale (VADTRS) consists of two sections: behaviours and performance.

The behaviours section assesses behavioural problems in a school setting and comprises 35 items divided into three subscales: Inattentive and Hyperactive/Impulsive (18 items), Oppositional-Defiant/Conduct (ODD, 10 items), and Anxiety and Depression (7 items). All behaviours are rated on a 4-point scale (0 = Never, 1 = Occasionally, 2 = Often, 3 = Very often) to indicate the frequency of these behaviours.

The performance section assesses related impairments in a school setting and includes two subscales: Academic performance (3 items) and Classroom Behavioural performance (5 items). Performance is rated on a 5-point scale (1, 2 = Problematic, 3 = Average, 4, 5 = Above average). This rating scale is user-friendly and serves as a valuable tool for clinical diagnosis by

collecting information within a school environment (Wolraich, 2003). You can find the items of VADTRS in Appendix 9.

Copeland Symptom Checklist for Attention Deficit Disorders

Edna D. Copeland (1987) developed this behavioural checklist to assist healthcare professionals and educators in assessing whether a child or adolescent exhibits symptoms of Attention Deficit Hyperactivity Disorder, the degree to which they do so, and which areas of functioning are most profoundly affected. This is a broad-band rating scale encompassing ten subscales, each of which corresponds to ten areas related to ADHD symptoms and impairments.

The ten subscales are as follows: Inattention/Distractibility, Impulsivity, Activity Level Problems, Noncompliance, Attention-Getting Behaviour, Immaturity, Poor Achievement/Cognitive & Visual-Motor Problems, Emotional Difficulties, Poor Peer Relations, and Family Interaction Problems.

The scale consists of 75 items, and each item is rated on a 4-point scale (0 = Not at all, 1 = Just a little, 2 = Pretty much, 3 = Very much) to indicate the severity level of the symptoms. You can find the items listed in Appendix 10.

Swanson, Nolan, and Pelham-IV Questionnaire (SNAP-IV)

The Swanson, Nolan, and Pelham-IV Questionnaire (SNAP-IV) is a rating scale designed to assess ADHD and related impairments and comorbidities in children aged 5 to 11 (Swanson, 1992). It was the first of several scales to adapt DSM symptoms of ADHD into a rating scale. To enhance inter-rater agreement and reduce discrepancies between parents and teachers arising from the use of different versions of rating scales, this scale employs a uniform format for completion by both parents and teachers.

The SNAP-IV is a broad-band scale with 90 items in the full version, evaluating a wide spectrum of ADHD impairments and associated comorbidities, including Oppositional Defiant Disorder (ODD), Conduct Disorder, Anxiety Disorder, and other externalizing and internalizing issues.

In addition to the DSM-IV symptoms for ADHD and ODD, the SNAP-IV incorporates items from other widely used rating scales to gauge the severity of the disorder. These include the Conners Index Questionnaire (Conners, 1968), the IOWA Conners Questionnaire (Loney & Milich, 1985), and the Swanson, Kotkin, Agler, MyInn, and Pelham (SKAMP) Rating Scale. Items from the Conners Index Questionnaire were selected based on the highest loading index across multiple factors of the Conners Questionnaire, representing a general index of childhood problems. Items from IOWA assess inattention/overactivity and aggression/defiance. The SNAP-IV-C also includes the 10 items of the SKAMP Rating Scale, which illustrate classroom manifestations of inattention, hyperactivity, and impulsivity and may be used to estimate classroom-related impairment.

The SNAP-IV employs a 4-point scale (Not at All = 0, Just A Little = 1, Quite a Bit = 2, and Very Much = 3) to assess the frequency of symptoms. It has been widely used in ADHD research, examining the effectiveness of ADHD treatments, investigating familial clustering of ADHD and other disruptive behaviours, and exploring the differential effects of ADHD subtypes on intelligence, depression, and other externalizing behaviours (MTA Cooperative Group, 1999; Willcutt et al., 1999).

Given the SNAP-IV's comprehensive coverage of impairments and disorders linked to ADHD, there are also shorter versions tailored for more specific assessments. The SNAP-IV-26, for

example, comprises 26 items and was adapted from the original SNAP-IV for the Multimodal Treatment Study of Children with ADHD (MTA) conducted in 1999. The SNAP-IV-26 screens for nine symptoms of ADHD hyperactive-impulsive type, nine symptoms of ADHD inattentive type, and eight symptoms of Oppositional Defiant Disorder as defined in the DSM-IV. Risk scores are generated by summing and calculating average scores for each symptom cluster and a combined score for the hyperactive-impulsive and inattentive clusters. You can find the items listed in Appendix 11.

The SKAMP 10 items (see in *Table 11*) subscale assesses the impairment of ADHD in the classroom. As SKAMP has been developed for the school setting to allow teachers to indicate their concerns about children with ADHD, it has been widely used in studies requiring repeated administration of scale to assess the effectiveness of treatments in the classroom setting.

Table 11 The SKAMP items

Items
Has difficulty getting started on classroom assignments
Has difficulty staying on task for an entire classroom period
Has problems in completion of work on classroom assignments
Has problems in accuracy or neatness of written work in the classroom
Has difficulty attending to a group classroom activity or discussion
Has difficulty making transitions to the next topic or classroom period
Has problems in interactions with peers in the classroom
Has problems in interactions with staff (teacher or aide)
Has difficulty remaining quiet according to classroom rules
Has difficulty staying seated according to classroom rules

In general, the narrow-band rating scales are designed to align with the DSM-IV diagnostic criteria for ADHD. These 18 symptoms constitute the items for rating scales. In contrast, broad-band rating scales are intended to assess not only the 18 core symptoms of ADHD but also the disorder's severity, related impairments, and comorbidities. While these reviewed rating scales do not encompass all those available for ADHD, they are among the most commonly used tools for gathering information from teachers within a school setting.

To gain a deeper understanding of teachers' perceptions of ADHD behaviours in the classroom, it was crucial to review these teacher rating scales. Furthermore, the items included in these rating scales provide comprehensive descriptions of the behaviours exhibited by children with ADHD in various contexts. These descriptions prove invaluable in creating vignettes to identify and gauge teachers' perceptions of ADHD behaviours in their classrooms. To extract the details and contexts from the rating scales, the study adhered to the DSM-V ADHD symptoms and categorized the items from these rating scales into 18 symptoms. The categorised items are shown in *Table 12*.

Table 12 Extracted behaviours of ADHD from rating scale review

Criteria	Extracted Behaviours
Inattentive	**Often fails to give close attention to details or makes careless mistakes in schoolwork, work, or during other activities (e. g. overlooks or misses details, work is inaccurate).** ● Rushes through assignments with little or no regard for accuracy or quality of work ● Loses place when reading (e.g., leaves out words, lines, or sentences when reading, etc.) ● Omits, adds, substitutes, or reverses letters, words, or sounds when reading ● Fails to copy letters, words, sentences, and numbers from a textbook, chalkboard, etc. ● Omits, adds, or substitutes words when writing ● Completes assignments with little or no regard for neatness (e.g., rushes through tasks, does not care to do well, etc.) ● Has problems in accuracy or neatness of written work in the classroom

Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).

- Only pays attention to things he/she is really interested in
- Has difficulty concentrating (e.g., staying on an assigned task, following a conversation, etc.)
- Daydreaming
- Engages in much activity, but accomplishes little
- Enthusiastic beginnings, but poor endings
- Has difficulty staying on task for an entire classroom period

Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).

- Does not listen to what other students are saying
- Does not hear all of what is said (e.g., misses word endings, misses keywords such as "do not," etc.)
- Does not direct attention or fails to maintain attention to important sounds in the immediate environment (e.g., teacher directions, public address system, etc.)
- Is unsuccessful in activities requiring listening (e.g., games, following oral directions, etc.)
- Requires eye contact in order to listen successfully (e.g., one-to-one situation, etc.)
- Does not listen to or follow verbal directions

Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily distracted).

- Fails to finish things he/she starts
- Sees tasks through to the end, good attention span
- Fails to complete homework assignments and return them to school
- Does not perform or complete classroom assignments during class time (e.g., does not perform the assignment or does not use the time provided, will go on to another assignment before completing the first, etc.)
- Does not read or follow written directions (e.g., instructions for homework, assignment directions, etc.)
- Begins assignments before receiving directions or instructions or does not follow directions or instructions
- Has problems in completion of work on classroom assignments
- Difficulty following the rules of games and social interactions

Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).

- Nervous or clingy in new situations, easily loses confidence
- Needs oral questions and directions frequently repeated (e.g.,

student says, “I don’t understand,” needs constant reminders, etc.)
- Attends more successfully when close to the source of sound (e.g., when seated close to the teacher, etc.)
- Fails to demonstrate short-term memory skills (e.g., does not remember two- or three-step directions, does not remember materials needed for a task, etc.)
- Fails to remember sequences (e.g., events in a daily routine, days of the week, months of the year, etc.)
- Is disorganized to the point of not having necessary materials, losing materials, failing to complete assignments, failing to follow the steps of the assignment in order, etc.
- Fails to perform assignments independently (e.g., continually asks for assistance or reassurance; does not begin, work on, or complete assignments without assistance; etc.)
- Fails to make appropriate use of study time (e.g., does not read, study, work on assignments, etc.)
- Fails to follow necessary steps in math problems (e.g., does step in the wrong order, omits a step, etc.)
- Changes from one activity to another without finishing the first, without putting things away, before it is time to move on, etc.
- Has difficulty making transitions to the next topic or classroom period

Often avoids or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).
- If a task is difficult, the child gets frustrated and quit.
- Lacks interest in schoolwork
- Does not prepare for school assignments (e.g., does not study for tests or quizzes, does not read the assigned material, etc.)
- Does not remain on-task (e.g., is more interested in other activities, sits and does nothing, etc.)
- Has difficulty getting started on classroom assignments
- Low frustration tolerance

Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).

Is often easily distracted by extraneous stimuli (e.g., for older adolescents and adults may include unrelated thoughts).
- Easily distracted, concentration wanders
- Is easily distracted by other activities in the classroom, other students, the teacher, etc.
- Engages in activities which do not relate to learning

	Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments). ● Forgets to hand in homework on time. ● Neglect or forget regular chores or tasks ● Forgets things he/she has already learned
Hyperactive	**Often fidgets with or taps hands or squirms in seat.** ● Restless in the "squirmy" sense ● Moves about while seated, fidgets, squirms, etc. ● Handles objects (e.g., twirls pencils, plays with things in desk, spins rulers on pencils, clicks ballpoint pens, repeatedly sharpens pencils, etc.) ● Engages in inappropriate behaviours while seated (e.g., tips chair or desk, puts feet on the desk, touches others as they walk by, taps and makes noises, etc.) ● Engages in nervous habits (e.g., bites fingernails, twirls hair, chews inside of the cheek, chews pencils or pens, spins or twirls objects, etc.) **Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplaces, or in other situations that require remaining in place).** ● Leaves seat without permission ● Moves about unnecessarily (e.g., leaves the seat, walks around the classroom, rocks, shakes head, etc.) ● The child is accident-prone. **Often runs about or climbs in situations where it is inappropriate (e.g., in adolescents or adults, may be limited to feeling restless).** ● Does not thinks things out before acting ● Demonstrates inappropriate behavior when moving with a group (e.g., fails to stay in line, runs, pushes, etc.) **Often unable to play or engage in leisure activities quietly;** ● Does not adjust behavior to the expectations of different situations (e.g., gets excited at recess and does not settle down, etc.) ● Becomes overexcited (e.g., loses control in group activities, becomes loud, etc.) ● Hums and makes other odd noises **Is often "on the go" acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for an extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).** ● Appears restless (e.g., shifts position in seat, paces about, etc.) ● Hops, skips and jumps when moving from one place to

	another instead of walking ● Excessively shifts from one activity to another ● Unable to stop when he/she is told to stop ● Unable to calm down when he or she is excited or all wound up **Often talks excessively**; ● Has difficulty remaining quiet according to classroom rules ● Talks to others during quiet activity periods ● Talks beyond what is expected or at inappropriate times
Impulsive	**Often blurts out answers before questions have been completed (e.g., completes people's sentences; cannot wait for turn in conversation).** ● Can not wait for an answer and jump to something else, e.g., new questions ● Blurts out answers without being called on ● Frequently interrupts conversations talks out turn **Often has difficulty awaiting turn (e.g., while waiting in line).** Does not wait for his/her turn in activities or games Is impulsive (e.g., reacts immediately to situations without thinking, is impatient, fails to wait for a turn or for assistance from an instructor, etc.) Does not wait appropriately for assistance from the instructor Difficulty in group situations which require patience and taking turns Has difficulty attending to a group classroom activity or discussion **Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities. May start using other people's things without asking, or receiving permission; for adolescents and adults, may intrude into, or take over what others are doing).** ● Grabs things away from others ● Interrupts the teacher (e.g., begins talking while the teacher is talking, goes up to the teacher while the teacher is working with other students, etc.) ● Interrupts other students (e.g., talks while they are talking, makes noises, laughs, etc ● Bothers other students who are trying to work, listen, etc. ● Makes unnecessary comments or noises in the classroom (e.g., talks to other students without permission, interrupts, makes fun of others, hums, taps, burps, etc.) ● Makes unnecessary physical contact with others (e.g., touches, hugs, etc.) ● Butts into games or activities even when he or she hasn’t been invited

7.2.6 The Final Draft of Cartoon Scripts

Through a comprehensive multi-method approach, incorporating classroom observation, teacher interviews, and a review of rating scales, the study successfully constructed an extensive ADHD behavioural bank, enriched with detailed contexts. Utilising this valuable resource, the researcher crafted a series of scripts, adeptly depicting children's ADHD behaviours within the classroom environment. Several rounds of rigorous review and revision, guided by input from PhD supervisors and school teachers from both countries, resulted in the final draft of the cartoon scripts. The final scripts of the six cartoon animations are in *Table 13* to *Table 18*.

Table 13 Cartoon Can't wait script

Cartoon title: Can’t Wait	
Symptom description: Often has difficulty waiting his or her turn (e.g., while waiting in line).	
Scenario: The kids are sitting at their desks. The teacher comes in with a plate full of sweets (look attractive and colourful).	
Child A	Leaving the seat, going to the teacher, and tapping the teacher’s body to ask for sweets. (The teacher points the seat, and the child gets back to the seat)
Child B	Sitting in the seat but shouting, “I want sweets(糖), I want sweets(糖)”. (“糖” is the Chinese character for sweet. I need two different languages here. No sound, just cartoon dialogue)
Child C	Looking at the sweets and sitting appropriately, and waiting for his/her turn
Child D	Leaving the seat and grabbing the sweets from the plate
Child E	Leaving the seat and hanging around with the plate.

Table 14 Cartoon Fidgety script

Cartoon title: Fidgety	
Symptom description: Often fidgets with or taps hands or feet, or squirms in their seat.	
Scenario: The cartoon should be based in a classroom with children seated at desks in positions so that their faces, arms and legs can be clearly seen. The teacher is standing and looking at the group of children as if speaking to them.	
Child A	Looking at the teacher but leaning on the desk, moving his legs, squirming in his seat (Action loop).
Child B	Looking at the teacher and apparently concentrating but fiddling with a pencil and tapping feet (Action loop).
Child C	Looking at the teacher but moving chair (making noise) (Action loop).
Child D	Looking at the teacher but moving chair (making noise) (Action loop).
Child E	Looking at the teacher but fiddling with the desk mate's hair, twirling it around in his/her fingers.

Table 15 Cartoon Inattentive script

Cartoon title: Inattentive	
Symptom description: Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).	
Scenario: The teacher is teaching the alphabet (A, B, C) on the blackboard.	
Scene 1: Child A, B, C, D, and E sit in the classroom	
Scene 2: The teacher is talking, and the background has a blackboard with A, B, and C characters on it.	
Child A	Not Looking at the teacher and drawing on a piece of paper (Zoom in. A hand is drawing something irrelevant to the lesson on a piece of paper).
Child B	looking at the teacher for a few seconds and then tearing a piece of paper for a few seconds (Action loop: looking-tearing-looking-tearing).
Child C	Looking at the teacher for a few seconds and then begins looking around and daydreaming.
Child D	looking at the teacher for one or two seconds and then taking snacks out of pocket and putting them into the mouth (Action loop: looking-eating-looking-eating).
Child E	Looking at the teacher and concentrating.

Table 16 Cartoon Disorganised

Cartoon title: Disorganised	
Symptom description: Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).	
Scenario: A teacher is standing in front of the classroom and waiting for all the kids to be ready for the lesson. She expects all children sit still with their books and stationery on the table.	
Child A	Walking into the classroom late with his school bag. (walking in at the end of the cartoon).
Child B	Sitting at the table and looking at the teacher, but the table is in a mess (books and stationery are in a mess) (Zoom in books and pencils are in a mess on the table).
Child C	Sitting at the table and looking at the teacher, the table is tidy (books and stationery are tidy on the table) (Zoom in books and pencils are tidy on the table).
Child D	Sitting at the table and looking for something on the table and under the table.
Child E	Leaning on the table and no books and stationery on the table.

Table 17 Cartoon Distracted script

Cartoon title: Distracted	
Symptom description: Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).	
Scenario: The teacher is teaching, and all the kids are listening. Then a butterfly flies into the classroom.	
Scene 1: A teacher is teaching A, B, and C in front of a blackboard. We can see A, B, and C on the blackboard, and the teacher is talking.	
Scene 2: Classroom, five children are sitting on the table and looking at the teacher. A Butterfly flies from the top left corner of the classroom.	
Child A	Walking into the classroom late with his school bag. (walking in at the end of the cartoon). Looking at the butterfly and looking back to the teacher again. When the butterfly flies in front of Child A's face, Child A's Eyes open widely and the head moves, then look at the teacher.
Child B	Looking at the butterfly, and the eyes are following the butterfly all the time. When Child B notices the butterfly, Child B's eyes follow the butterfly with the head moving up and down and turning.
Child C	Playing with the butterfly but staying in the seat. When Child C notices the butterfly, Child C is trying to reach the butterfly and even stands up. After the butterfly flies out of his reach, Child C sits down, but eyes follow the butterfly.
Child D	Shouting "butterfly, butterfly". When Child D notices the butterfly, Child D shouts out and points to the butterfly. Repeated action.
Child E	Looking at the butterfly and leaving the seat to chase the butterfly. After Child D shouts, Child E notices the butterfly and leaves the seat to chase the butterfly. Body and arm movements of catching the butterfly.

Table 18 Cartoon Disruptive script

Cartoon title: Disruptive	
Symptom description: Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission)	
Scenario: The teacher is teaching how to write character A on the blackboard.	
Scene 1: The teacher is writing character A on the blackboard. After she has finished, she gives a sign for the children to practice character A on a piece of paper.	
Scene 2: Students are sitting at the table with a pencil in hand and looking at the teacher and starting writing. Child B has no pencil in hand.	
Child A, Child B, Child C, Child D and Child E behaviours happen one by one. NOT at the same time.	
Child A	Sitting and writing
Child B	Sitting but has no pencil, taking a pencil from Child A without asking. Child A looks annoyed and picks up another pencil and continues writing.
Child C	Sitting and writing, then moving body close to Child B and looking at Child B's writing and tickling Child B. Child B looks annoyed but continues writing.
Child D	Sitting and writing, looking at Child C's writing and drawing lines on Child C's writing.
Child E	Butting into the teacher's teaching and holding his writing and waving it in the air and trying to show his writing to the teacher. The teacher shows a stop sign to Child E. However, Child E was waving his writing to the teacher again and trying to get her attention.

7.2.7 Cartoon Animations Production

The production of cartoon animations is a highly intricate and professional undertaking, encompassing various stages and meticulous attention to detail. The process began by assembling a team of skilled animators and designing the characters and background settings. Storyboard drawing followed, laying the foundation for the animation sequence. Working in collaboration with the animators, the researcher transformed the carefully crafted cartoon

scripts into captivating and authentic portrayals of children's ADHD behaviours within the classroom setting. The result was a series of captivating cartoon animations, serving as a valuable resource for exploring rater bias and facilitating cross-cultural comparisons in ADHD assessments.

The Ways of Making Cartoon Animation

This process involved sourcing the animators, storyboard drawing, character design, animation, and revisions. There are several methods for creating animation. The first approach is using animation software, which offers flexibility and cost-effectiveness. Various open-source and free animation software options, such as Pencil 2D, Plastic Animation Paper, Synfig Studio, Creatoon, Stykz, Toon Boom Animation Pro 3, and more, are available. These software tools are not only powerful for professional animators but also user-friendly for beginners.

With this software, you can create scenes and characters from sketches using tools like pens, pencils, and brushes integrated into the software. Alternatively, you can select scenes and characters from the software's material library. These programs enable creators to animate characters frame by frame into 2D animations, incorporating features like bitmap, vector lines, and advanced functions such as onion skinning. While these software options are free and relatively beginner-friendly, they do require basic skills in drawing, painting, acting, modelling, visualization, and a good sense of timing and spacing. After attempting to develop a cartoon (titled 'Can't Wait') using one of these software packages, I found that I couldn't achieve the desired quality due to my lack of these essential skills.

The second method involves using animated video templates, which can be found on many websites. These templates encompass scenes, characters, and a wide range of effects, offering finished or partially completed 2D and 3D animations for download. Creators can obtain the

components they need and combine them to craft their cartoons. It's akin to assembling a car using parts purchased online. The advantages include relative affordability, with some templates available for free, and a vast array of choices. However, the downside is the overwhelming number of options; much like online shopping, it can be challenging to find precisely what a creator needs to develop a series of cartoons.

The third approach is engaging a professional animation company. The benefit here is the ability to obtain precisely what you desire, with their completed works typically of high quality. The drawback is the cost, as these companies prioritise projects based on their budget. Consequently, they excel in creating commercial animations, such as advertisements, television shows, and films. After contacting several animation companies, none were willing to undertake such a small budget and project. Therefore, it is not my option for developing my cartoons.

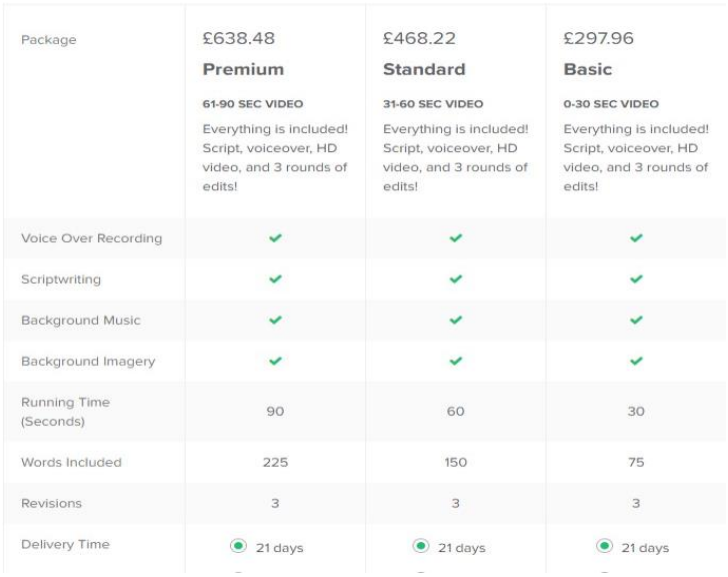
Choosing Freelance Animators

The final option available to me was to hire freelance animators. There are several websites for freelance animators, such as Upwork, PeoplePerHour, Fiverr, and Toptal. Freelance animators can be found worldwide, and their profiles typically showcase their specialities, completed projects, customer reviews, average response times, and recent delivery times.

To engage freelance animators on these platforms, there are two primary methods. One approach involves posting the project as a job and providing clear descriptions of your expectations for the job and the budget you've allocated. Freelance animators interested in the project will then reach out to you through the website, allowing you to select the animator you believe is best suited for the job.

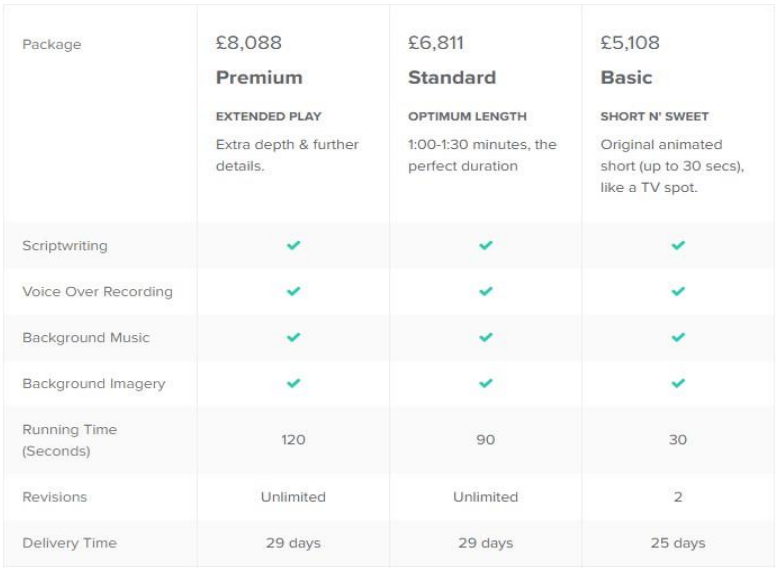
Alternatively, you can explore the profiles of freelance animators and directly contact those who pique your interest. The cost of freelance cartoon animation can vary significantly, depending on the complexity of the work, the desired level of quality, and the project's timeline. These are two examples of the charges of freelance animators in *Figure 6* and *Figure 7.*

Figure 6 Freelance animator charges example 1



Package	£638.48 **Premium** 61-90 SEC VIDEO Everything is included! Script, voiceover, HD video, and 3 rounds of edits!	£468.22 **Standard** 31-60 SEC VIDEO Everything is included! Script, voiceover, HD video, and 3 rounds of edits!	£297.96 **Basic** 0-30 SEC VIDEO Everything is included! Script, voiceover, HD video, and 3 rounds of edits!
Voice Over Recording	✔	✔	✔
Scriptwriting	✔	✔	✔
Background Music	✔	✔	✔
Background Imagery	✔	✔	✔
Running Time (Seconds)	90	60	30
Words Included	225	150	75
Revisions	3	3	3
Delivery Time	21 days	21 days	21 days

Figure 7 Freelance animator charges example 2



Package	£8,088 **Premium** EXTENDED PLAY Extra depth & further details.	£6,811 **Standard** OPTIMUM LENGTH 1:00-1:30 minutes, the perfect duration	£5,108 **Basic** SHORT N' SWEET Original animated short (up to 30 secs), like a TV spot.
Scriptwriting	✔	✔	✔
Voice Over Recording	✔	✔	✔
Background Music	✔	✔	✔
Background Imagery	✔	✔	✔
Running Time (Seconds)	120	90	30
Revisions	Unlimited	Unlimited	2
Delivery Time	29 days	29 days	25 days

I opted for the second approach and identified two freelance animators. One hails from India, charging £200 for a single cartoon animation (30 seconds in length), which includes

background music, background imagery, two revisions, and a seven-day delivery timeframe. The other is based in Ukraine, working with a team. Their pricing structure involves £350 for the initial cartoon animation and £250 for subsequent ones (each 30 seconds long), encompassing background music, background imagery, unlimited revisions, and a 14-day delivery period. Both animators boast positive reviews and rapid delivery times. I decided to collaborate with the Indian animator and provided him with the script for 'Fidgety.'

However, our communication encountered difficulties. After receiving all the project requirements and the cartoon script, the Indian animator did not seek clarification or pose any questions. Although he delivered the animation swiftly, some character actions deviated from the script, and the character designs fell short of the specified criteria. I had to prompt him to re-evaluate the requirements and script to ensure accurate comprehension. Following two revisions, the animation met the required standards, but the animator declined further alterations.

I was dissatisfied with both his work and his attitude. Consequently, I sent the 'Fidgety' script to the Ukrainian animation team. Their project manager conducted a Skype meeting with me before commencing work. We meticulously reviewed all script details and discussed the objectives for the series of cartoons. Their initial cartoon delivery was of exceptional quality, leading to their selection for the project. *Figures 8* and *9* showcase animations from Indian and Ukrainian animators, with the Ukrainian animator delivering superior.

Figure 8 The animation from the Indian animator

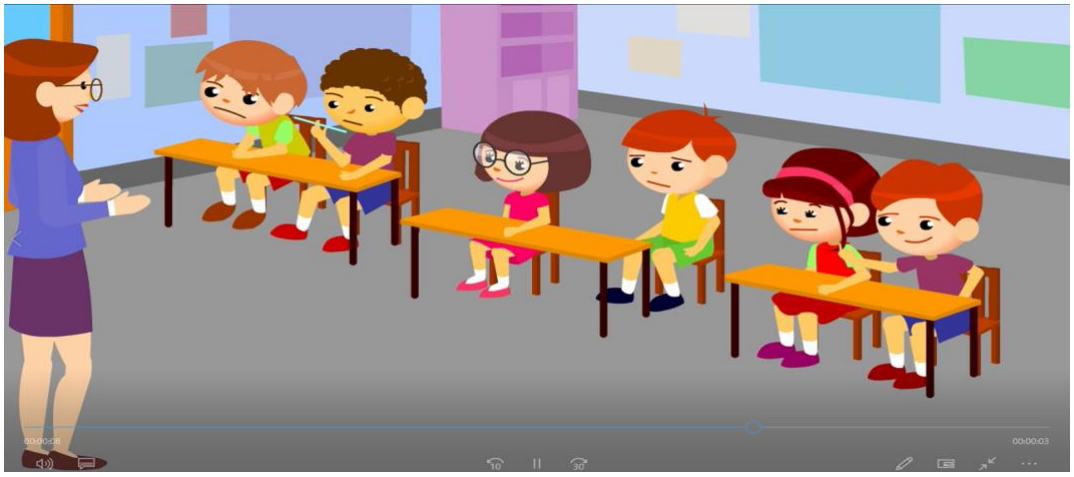


Figure 9 The animation from the Ukrainian animator

When the scripts reach their final draft, work commences on the storyboard. A storyboard serves as a visual interpretation of the script, comprising small thumbnail drawings that depict every scene outlined in the script. It utilises a range of symbols and sequences of images to indicate actions and camera movements (such as transitioning from wide to full to close shots and back) to map out how the story unfolds visually. However, it doesn't directly translate the script into images. Instead, the storyboard sets up the scenes effectively, introducing dramatic camera movements and transitions where necessary to enhance the storytelling. Once the

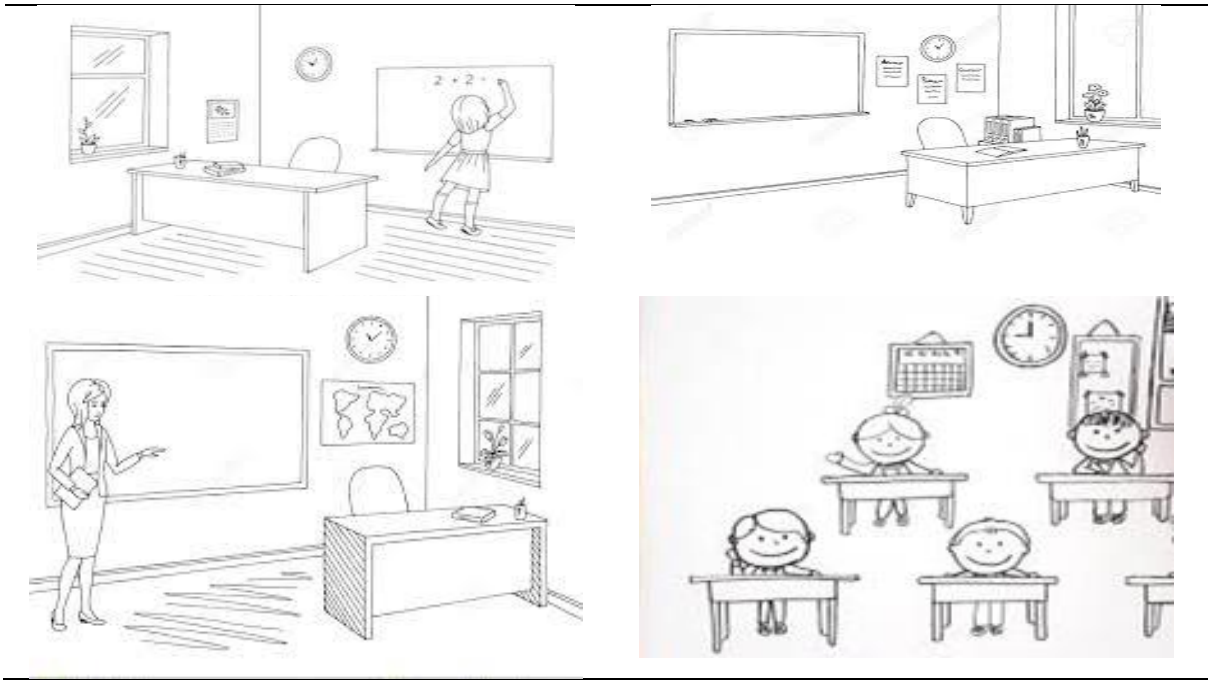
storyboard is ready for production, the artwork's design phase kicks off. This encompasses background design (sketches of the interior and exterior environments featured in the cartoon), character design (sketches of the characters), and designs for any other elements that might appear in the cartoon.

Background Design

The background design draws inspiration from the descriptions of the scenarios outlined in the six scripts. Since all the stories are set in a classroom, the background design encompasses sketches of a classroom from various camera angles. The storyboard artist initially crafted sketches for the key scenes in this series of cartoons, relying on their interpretation of the scripts. Subsequently, we engaged in discussions regarding the timeline and camera angles for each cartoon. *Table 19* provides an overview of the scenarios and background sketches.

Table 19 The scenarios and background sketches example

Scenarios
Cartoon 1 Can't wait
Scenario: The kids are sitting on their desks. The teacher comes in with a plate full of sweets (look attractive and colourful).
Cartoon 2 Fidgety
Scenario: The cartoon should be based in a classroom with children seated at desks in positions so that their faces, arms and legs can be clearly seen. The teacher is standing, looking at the group of children as if speaking to them.
Cartoon 3 Inattentive
Scenario: The teacher is teaching the alphabet (A, B, C) on the blackboard.
Cartoon 4 Disorganized
Scenario: A teacher is standing in front of the classroom and waiting for all the kids to be ready for the lesson. She expects that all children sit still with their books and stationery on the table.
Cartoon 5 Distractive
Scenario: The teacher is teaching, and all the kids are listening. Then a butterfly flies into the classroom.
Cartoon 6 Disruptive
Scenario: The teacher is teaching how to write character A on the blackboard.
Background Sketches



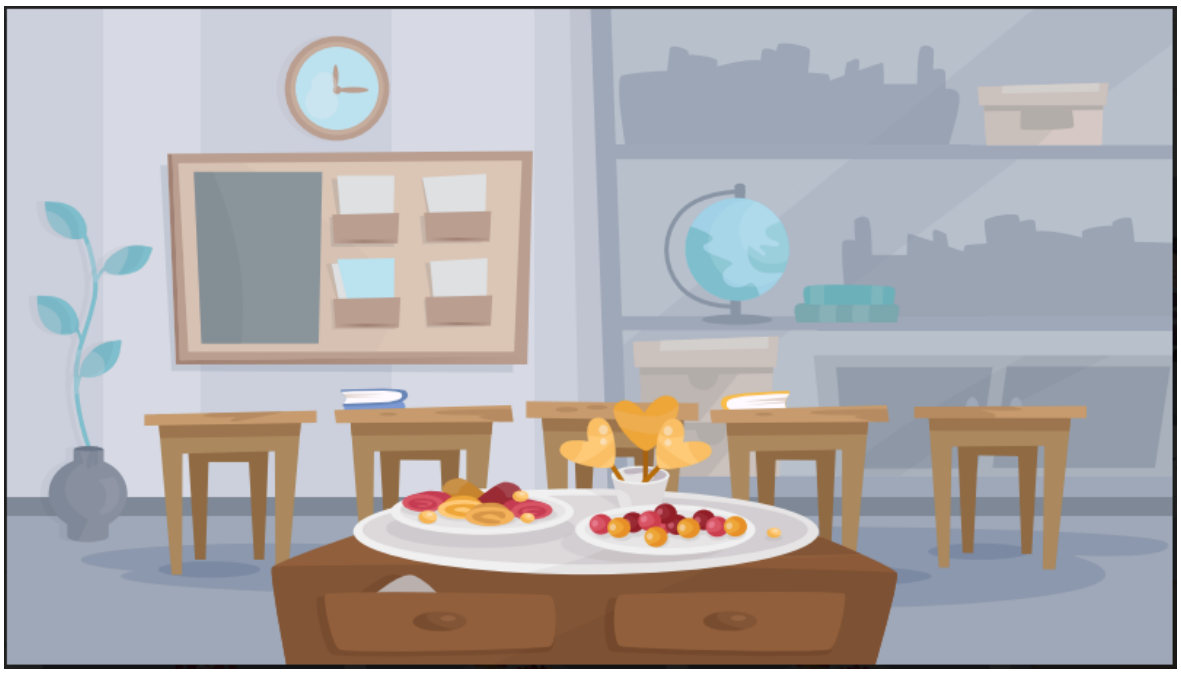
Character design

There are six characters in the cartoons. One of these characters is a teacher, while the remaining characters are children. Notably, the children in the cartoons lack distinct characteristics associated with any particular country or gender reference. The design of these characters encompasses considerations like skin tone, facial features, and clothing. Their appearance should resemble that of 6 to 7-year-olds, with each child assigned an identifying label (A, B, C, D, and E) for reference and evaluation. As an example, one of my friends, who has experience drawing cartoons for children, designed two characters for the cartoon, as illustrated in Figure 10. However, these characters bore a closer resemblance to European children. Subsequently, the storyboard artist developed several character designs. I presented these characters to both English and Chinese teachers for assessment. Ultimately, I selected characters that did not conform to any specific cultural or gender norms, as depicted in *Figure 11*. The design of the classroom environment is depicted in *Figure 12*.

Figure 10 The firs design of characters

Figure 11 The final design of characters

Figure 12 The classroom design



Storyboard

A storyboard consists of a series of drawings based on scripts, serving as a visual guide throughout the animation production process. It outlines the sequence of scenes necessary to convey the story. Each drawing provides insight into what will be both seen and heard within the scenes. The storyboard incorporates early concepts of camera staging, transitions, potential visual effects to enhance each shot, audio notes, key character poses, and scene events. In essence, it encompasses various elements essential for the team to gain a comprehensive understanding of the final animation's appearance and atmosphere. Once the storyboard undergoes revisions and no further changes are required, the animation production proceeds to the next stage. *Table 20* shows the storyboard of cartoon can't wait. The script was translated into Ukrainian language for the animators to understand.

Table 20 The storyboard of cartoon Can't wait

Поведение детей

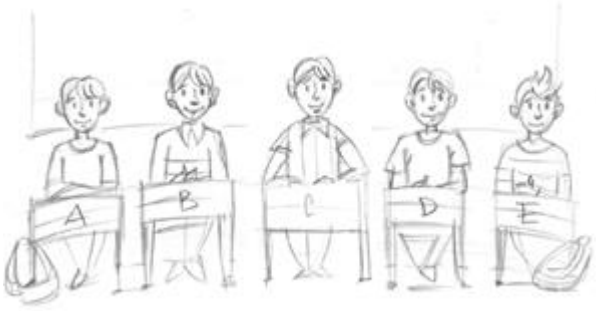
Цель: Создать серию анимационных роликов для учителей, что позволит мне, как исследователю, изучить, как они оценивают поведение детей.

У меня есть цель – сравнить оценку детей учителями в Китае и Англии, поэтому было бы хорошо, чтобы дети в анимации не имели принадлежности к какой-либо конкретной стране. Это касается их цвета кожи, формы лица/глаз, их одежды. Дети должны быть в возрасте 5-7 лет. То есть совсем маленькие, но и не взрослые.

Ролик 2: Сложности с ожиданием своей очереди (когда ребенок ждет стоит в очереди)

The kids are sitting on their desks in a round circle.

Дети сидят за своими партами в кругу.



The teacher comes in with a plate full of sweets (look attractive and colourful).

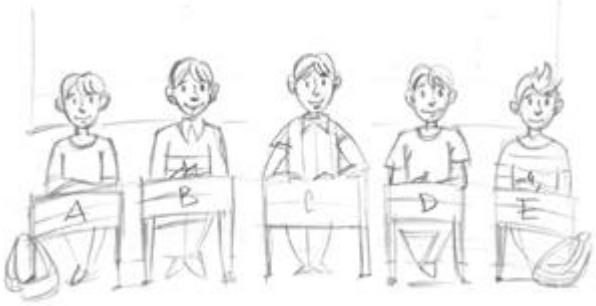
Учитель заходит в класс с тарелкой, полной сладостей (сладости выглядят ярко и аппетитно)

Child A: Leaving the seat, going to the teacher, tapping the teacher's body to ask for sweets.

Ребенок 1: Встает с сиденья, подходит к учителю, легко прикасаясь к нему, просит сладости.

The teacher points the seat and the child gets back to seat.

Учитель указывает на его место, и ребенок возвращается за парту.



Child B: Sitting in the seat but shouting “I want sweets(糖), I want sweets(糖)” .

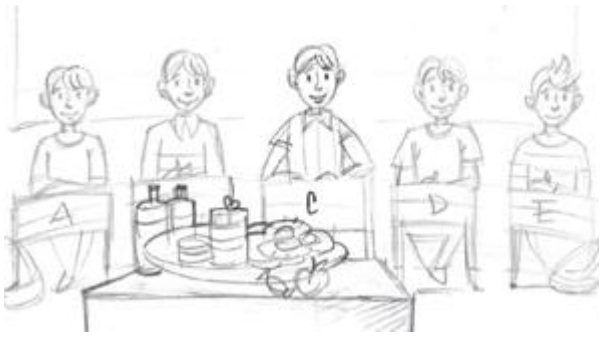
(“糖”Chinese character for sweet, I need two different languages here. No sound, just cartoon dialogue)

Ребенок 2: Сидит на своем месте, но при этом кричит «Я хочу сладости (糖), я хочу сладости»

(“ 糖 ” – это китайский иероглиф, обозначающий сладости. Нужно здесь сделать на 2ух языках. Без звука, только диалог в анимации)

Child C: Looking at the sweets and sitting appropriately and waiting for his/her turn.

Ребенок 3: Смотрит на сладости и прилежно сидит и ждет своей очереди



Child D: Leaving the seat and grabbing the sweets from the plate

Ребенок 4: Встает со своего места и хватает сладости с тарелки.

Child E: Leaving the seat and hanging around with the plate.

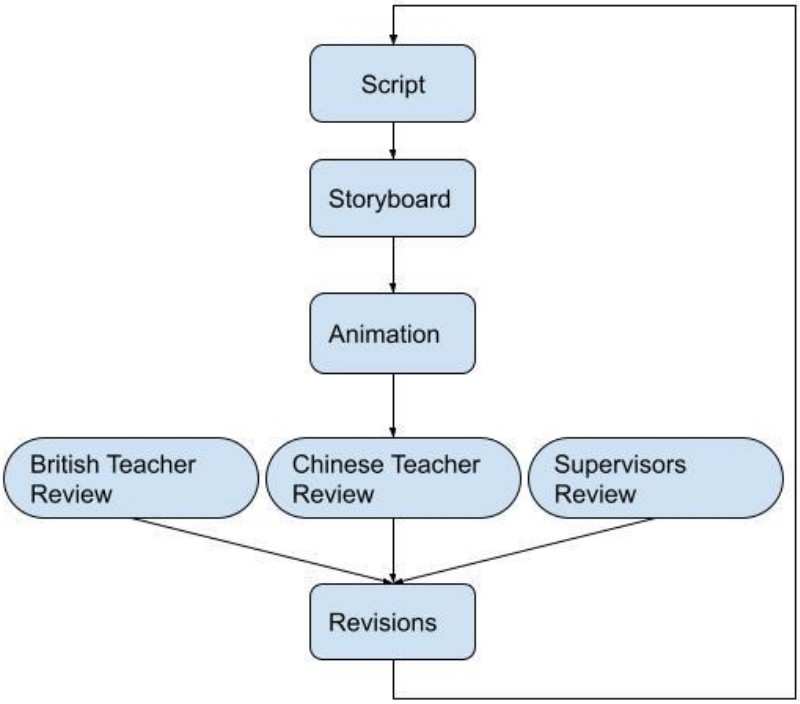
Ребенок 5: Встает со своего места и слоняется вокруг тарелки.

The Final Cartoon Animation

Once all the storyboards have been revised and finalised, the animators bring the illustrations to life using sophisticated software. They imbue characters and other on-screen elements with motion. Following the completion of the visual aspects, the animator incorporates background music for the cartoons. Typically, there are two or three rounds of revisions before the animated cartoons are considered finished. When scripts are transformed into final cartoon animations, occasional issues may arise. It's akin to adapting a novel into a film, where there's often a disparity between the scripts and the animations. This sometimes necessitates changes to either the scripts or the animations. These revisions follow a cyclic process aimed at bridging the gap between scripts and animations.

Figure 13 shows the revision procedure. The finished cartoons are in the link: Cartoon animations

Figure 13 The revision procedure



Cartoon animation is a team effort. The storyboard artist is the visual director. Her job is to transfer the words of the script into images and sometimes substantially change scenes to tell the story as creatively as possible, to be clear and maintain continuity, to increase emotion with dynamic visual storytelling, and to set the layout, background and timeline for the story. The storyboard artist works with the character model artist, the prop artist and the background artist to give them what they need to do their jobs well. The storyboard artist directly communicated with me to understand the project and my scripts. She collected all the information and put them on the board for the overseas artists who might be enlarging the panels for their layouts. The storyboard artist places doors, windows, furniture, and props in each room and plans the placement of characters and the camera. After the storyboard is done, the animators will use animation software to generate the final animation according to the storyboard.

Meanwhile, I was involved in all their creative work to ensure no important things were left out and also to be open-minded to take their suggestions to change the script. Although I wrote all the scripts, I do not have any experience in animation. Moreover, the animators will not see

the scripts the same way I see them. Their professional experiences and talents have made them see the scripts better than I do in animation. They have contributed to making the finished product better. The most challenging part of the animation is how to do it in a way that adds to a good story without complicating it. The animation must be done in a simple, clear way, especially when the finished product is for teachers to watch and rate the characters' behaviour. Therefore, the stories should be simple to make the point, and the ideas should be straightforward and communicated instantly and give the audience no more information than necessary. A confused audience will not give reliable and valid ratings. In order to achieve it, each cartoon animation has been reviewed by my supervisors, a Year 2 English teacher and a Year 1 Chinese teacher, to make sure the story and the actions are clear to them. With teamwork, the project has been visualized well, with each sequence attracting and holding the audience's attention and moving seamlessly into the next.

7.3 The Development of Teacher's Rating Scale of Cartoons Questionnaire

There are two questionnaires in the survey. One (Teacher's Rating Scale of Cartoons) is for measuring teachers' attitudes towards ADHD-type behaviours of the characters in the cartoons. The other (Teacher's Rating Scale of Children's Behaviour) is for rating the behaviours of children in the participants' teaching class.

7.3.1 Teacher's Rating Scale of Cartoons Design (TRSC)

TRSC (Teacher's Rating Scale of Cartoons) has adopted a semantic differential scale (SDS) to measure teachers' leniency towards ADHD behaviours. SDS is a rating scale developed by US psychologists Osgood, Suci, and Tannenbaum (1957). Initially, it aimed to establish a universal structure of meaning by bridging the gap between linguistic meaning and the philosophy of language using factor analysis. Factor analysis helped identify similarities and differences within and between participants' ratings, elucidating different dimensions of the connotative meaning of concepts, such as terms, objects, events, activities, and ideas. Participants were

required to rate concepts on a seven-point bipolar scale, using adjectival opposites, like negative and positive, honest and dishonest, based on their attitude or understanding within a semantic space. SDS finds extensive use in measuring social attitudes, particularly in linguistics and social psychology.

SDS offers several advantages. First, it highlights participants with extreme attitudes towards a concept, as they opt for the scale's extremities. Second, it signals when participants have not yet formed an opinion, as they choose a neutral position midway between the two extremes. Third, it compels subjects to concentrate on expected dimensions as categories are predefined (Agheyisi & Fishman, 1970). Fourth, it is relatively straightforward to implement. With a fixed bipolar scale encompassing negative and positive ends, participants find it relatively simple to grasp. The process isn't overly complicated; they merely need to make judgments and circle a number. In my research, TRSC aims to ascertain individual teachers' attitudes towards characters' behaviour in cartoon animations, making SDS well-suited for this purpose.

SDS comprises two main components: adjective selection and scale layout (Al-Hindawe, 1996). Several methods exist for selecting adjectives. One direct approach involves choosing complementary opposites, which can be a straightforward process for many adjectives, like honest and dishonest, happy and unhappy, lucky and unlucky, patient and impatient, and so forth. Some adjectives assume opposite meanings when negative morphemes are added to them. These pairs are employed to represent the bipolar scale's two opposing attitudes. However, situations may arise where paired adjectives aren't readily available or meaningful for the concept SDS aims to measure. Gradable antonyms present an alternative. Antonyms for certain concepts can encompass a wide array of adjectives. For instance, 'Healthy' may be opposed by 'Unhealthy,' 'Diseased,' 'Ill,' and 'Sick.' Nevertheless, there may be cases where choosing

appropriate adjectives isn't straightforward. For example, 'Unhealthy' doesn't necessarily equate to 'Diseased' or 'Sick,' just as 'Uneducated' doesn't necessarily mean 'Unskilled.' As discussed earlier, the choice between complementary and gradable antonyms hinges on the concept SDS seeks to measure.

However, this doesn't imply that only one type can be chosen. It's possible to employ a combination of complementary and gradable antonyms within the same scale. The determination of which adjectives are suitable hinges on their meaning's relevance to the concept SDS aims to measure and their meaningfulness within the specific context or culture of the study. This is because, based on the culture or context, people may attribute positive or negative characteristics to the same attitude object. For instance, 'ambitious' is positively perceived in Western culture but negatively in Eastern culture (Furuya-Nakajima & Vogt, 1990). Furthermore, some adjectives may hold positive connotations for males and negative for females, like 'tough.' Thus, words can possess varied interpretations for different individuals. Selecting adjectives that are both relevant to the study and resonate with the specific subject group is crucial.

Therefore, comprehending the concept SDS aims to measure is paramount. In TRSC (Teacher's Rating Scale of Cartoon), the scale endeavours to gauge teachers' attitudes towards cartoon characters' behaviours. An individual's attitude towards a particular behaviour pertains to the extent of their positive or negative evaluation or assessment of said behaviour (Stangor, Jhangiani & Tarry, 2014). In essence, it measures teachers' positive or negative appraisal of the cartoon characters' behaviours. Consequently, the scale necessitates a pair of positive and negative evaluation adjectives in English and Chinese that are unambiguous and free from misinterpretation.

Four pairs (A, B, C, and D) of adjectives are commonly employed to describe positive and negative attitudes toward the behaviours depicted in *Figure 14.* In terms of meaning, Group D possesses a broader and more explicit range than the other pairs in both English and Chinese.

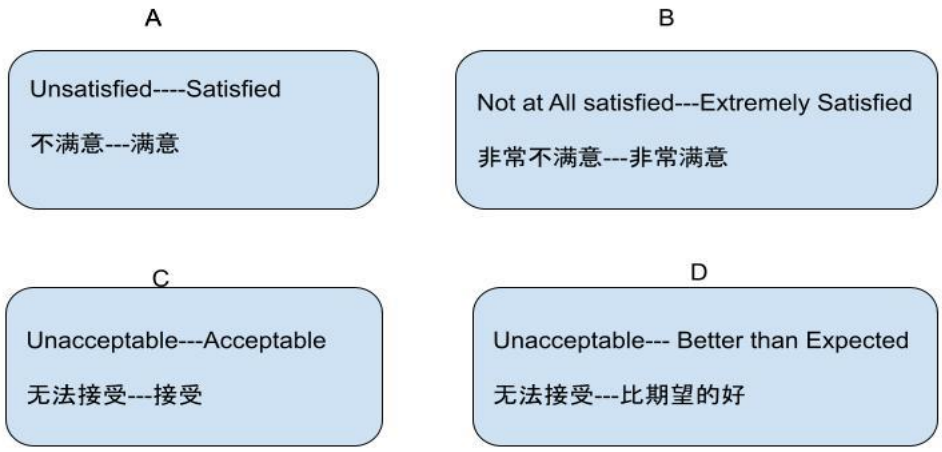
In Group A, the negative word 'Unsatisfied' conveys a sense of 'feeling that something is not enough or not good enough' (Cambridge Dictionary). It may suggest that something falls short of expectations but doesn't necessarily imply that it is bad. On the other hand, the positive word 'Satisfied' denotes 'being pleased because you have obtained what you desired or because something has occurred as you wished' (Cambridge Dictionary). It may suggest that something meets or even exceeds expectations, although it's not explicitly stated. Consequently, this pair of adjectives presents a relatively limited range of attitudes and some degree of interpretational ambiguity. Group B, while seemingly having a broader scope of meaning compared to Group A, still harbours a similar level of ambiguity.

In Group C, 'Unacceptable' signifies that something is 'too bad to be accepted, approved of, or allowed to continue,' suggesting a clear negative attitude towards 'bad'. However, 'Acceptable' implies that something is 'just good enough, but not very good.' It offers possible interpretations such as 'not bad,' 'good,' and 'not very good,' but it doesn't necessarily indicate that something exceeds expectations.

In contrast, Group D, where 'Better than Expected' implies results, profits, etc., that 'are higher than initially thought,' suggests that something exceeds expectations. Group D encompasses a broader spectrum of attitudes than the other groups, ranging from considering something as bad to surpassing expectations. It also exhibits less ambiguity in interpretation. Therefore,

Group D is the chosen option. Furthermore, the meanings of these words in the Chinese language have been verified to align with their meanings in the English language through testing the interpretation of this word pair with native English and Chinese speakers.

Figure 14 Four pairs of adjectives for semantic differential scale



The second component to consider is the layout of the SDS scale. When designing the layout, two key aspects must be taken into account: polarity and the number of scale points. In polarity design, the central issue revolves around how to arrange the positive and negative adjectives on the scale. The scale can adopt a consistent system with negative adjectives placed at the left end and positive adjectives at the right end. Research indicates that this layout is suitable for concepts where the adjective pairs have clear positive and negative meanings, such as 'Good' and 'Bad'. It helps respondents make more straightforward judgments and reduces mental effort (Alford & Strother, 1990). Alternatively, the scale can adopt a randomised system where positive and negative adjectives are allocated randomly for polarity. Research indicates that this layout is more fitting for concepts where there is no consensus on whether an adjective is positive or negative (Gallois & Callan, 1981; Romaine, 1980).

TRSC (Teacher's Rating Scale of Cartoons) assesses teachers' severity/leniency towards the characters' behaviour in cartoon animations. All behaviours are evaluated on the same scale,

ranging from 'Unacceptable' to 'Better than Expected', signifying an apparent positive and negative meaning. Therefore, TRSC has opted for a consistent system, placing the negative adjective on the left end and positive adjectives on the right. This layout facilitates respondent comprehension and judgment.

Regarding scale format design, there are discrete and continuous scales. The original SDS scale is a discrete seven-point scale (-3, 0, +3) (Osgood, Suci & Tannenbaum, 1957). With its widespread use, it has been adapted into different versions with five, six, and nine points. Logically, the more points a scale has, the finer the gradation it offers. However, having too many points can render differences between them meaningless, making it challenging for participants to provide such fine judgments. This can make the evaluation task tedious and counterproductive.

Another consideration when selecting the number of scale points is whether to use an odd or even number. The main distinction lies in whether the scale provides participants with a neutral option, allowing them to select a middle point. When participants choose the midpoint, it might signify that they hold neither a positive nor a negative attitude towards the object of evaluation, or they might be undecided or lack sufficient information to make a judgment. However, without follow-up questions, researchers remain uncertain about the precise meaning of these neutral points. The odd-numbered scale compels participants to make a clear positive or negative choice regarding the evaluation object. However, with a 6-point scale, it's unclear whether a participant circling '3' or '4' intends to express a more positive or negative sentiment or if they are simply neutral and forced to make a choice. This kind of scale can introduce errors in the research. Moreover, discrete semantic differential scales often suffer from shared

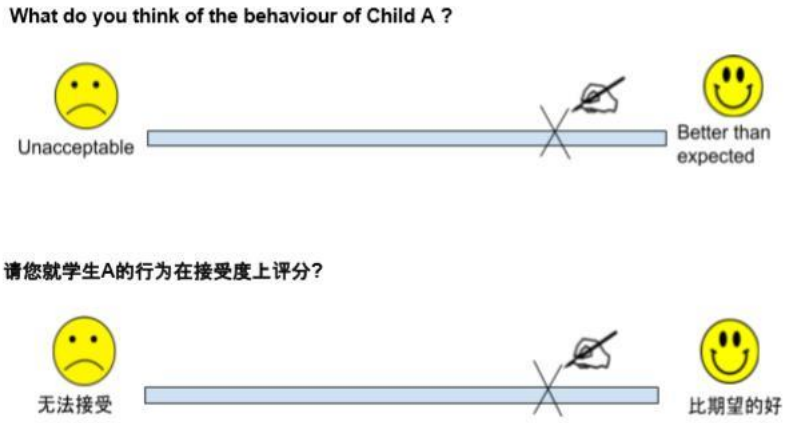
rankings and limited differentiation between aspects at the individual level (Yu, Keown & Jacobs, 1993).

Continuous scales, on the other hand, don't encounter these issues. Their primary advantage over discrete scales is that respondents aren't confined to predefined values; they can place a mark anywhere on the scale based on their opinions about the evaluation object. Once the ratings are collected, researchers divide the scale into categories and assign scores based on where the ratings fall. Alternatively, researchers measure the distance from the respondent's mark to the left end of the scale as the rating's value (Reips & Funke 2008). Research has demonstrated that continuous rating scales can provide the same aggregate information while affording greater discrimination at the individual level. Continuous semantic differential rating scales prove advantageous in studies where individual differences hold significance (Albaum, Best & Hawkins, 1981; Funke & Reips, 2012). This implies that continuous scales are more sensitive to detecting small differences among respondents without introducing meaningless points on the scale, which could perplex participants.

Moreover, continuous scales offer more flexibility in data analysis. In my research, TRSC (Teacher's Rating Scale of Cartoons) gauges teachers' attitudes towards characters' behaviours in cartoon animation. The ratings from this scale will be utilised to compare attitudes between two groups of English and Chinese teachers towards ADHD-type behaviours in the cartoon animations. Additionally, employing modern test theory, these ratings will be analysed using the Rasch model to determine individual teachers' acceptance of these behaviours (i.e., how tolerant the teachers are towards these behaviours). These individual differences in teacher acceptance are crucial for the data analysis in my research. Consequently, the continuous rating

scale is better suited to my research than the discrete rating scale. An example of the scale is depicted in *Figure 15.*

Figure 15 The rating scale for TRSC in English and Chinese



7.4 The development of Teacher's Rating Scale of Children's Behaviours (TRSCB)

The TRSCB (Children's Behaviour Rating) questionnaire requires teachers to rate six ADHD symptoms (Can't Wait, Fidget, Inattentive, Disorganized, Distractive, and Disruptive) of 10 randomly selected children from their teaching class. The rating scale employs a 5-point Likert scale (0 = Never, 1 = Occasionally, 2 = Often, 3 = Very Often, 4 = Almost Always) to assess how frequently these children exhibit the behaviours associated with the six types of ADHD symptoms depicted in the six cartoon animations.

Behaviour rating scales are commonly used in a school setting due to their ease of administration and quick results. The TRSCB (Children's Behaviour Rating) has adopted widely used behaviour rating scales completed by parents and teachers, which are extensively employed in clinical settings to obtain parent and teacher ratings regarding the frequency of each ADHD symptom based on DSM criteria. As in the earlier section of the ADHD Rating Scale Review, most rating scales employ 4-point or 5-point Likert scales to measure the frequency of behaviours.

TRSCB employs straightforward wording for items such as 'Can't Wait' rather than using items with descriptions similar to ADHD symptoms in DSM criteria, such as 'Often has difficulty waiting for his or her turn (e.g., while waiting in line)'. The 'Can't Wait' cartoon animation illustrates behaviours associated with this symptom within a specific context. Consequently, TRSCB includes only six items (Can't Wait, Fidgety, Inattentive, Disorganized, Distractive, and Disruptive), which are the titles of the six cartoon animations.

The behaviours of the six symptoms are portrayed in the six cartoon animations. Teachers were instructed to watch these animations and rate the behaviours of cartoon characters initially. Then, they were asked to rate 10 randomly selected children from their teaching class, assessing the children's behaviours in line with what they observed in the cartoon animations. Using the titles of cartoon animations as rating items for children's behaviour offers several advantages over using symptom descriptions.

First, it reduces the risk of bias arising from translating a heavily worded questionnaire into a different language, a common issue when employing assessments across diverse cultures (Yu, Keown & Jacobs, 1993). Second, using cartoon titles as rating items for children's behaviour provides teachers with contexts and behaviours from the cartoon animations as reference points when assessing the children in their class. As discussed in Chapter 3, there are numerous criticisms of the diagnostic criteria. One of these criticisms is that these diagnostic criteria merely describe ADHD symptoms without context. Consequently, when these criteria are transformed into items on ADHD behaviour rating scales, judgements based on these rating scales become subjective as they rely on how the raters interpret the criteria. Particularly in cross-cultural research, language differences are not the sole factor influencing ADHD

behaviour judgements; cultural differences also play a significant role. Therefore, employing a questionnaire that does not heavily depend on language and offers contextual and behavioural references for rating will mitigate potential biases. Examples of the questionnaire are displayed in *Figure 16.*

Figure 16 The rating scale of TRSCB in English and Chinese

Teacher's Rating Scale of children's behaviour

Rating example:

Could you please select 10 children randomly in your class and rate their behaviours according to the behaviours in the cartoons?

0 = Never 1= Occasionally 2= Often 3 = Very Often 4 = Almost Always

Name	Can't wait	Fidgety	inattentive	Disorganized	Distracted	Disruptive
James	0 1 2 ③ 4	0 ① 2 3 4	0 ① 2 3 4	0 1 2 ③ 4	0 1 ② 3 4	0 1 ② 3 4

儿童行为评估教师问卷

测评样例:

请您对您班级中随机挑选的10个学生的课堂行为根据卡通中儿童的行为进行评估

0 = 从不 1= 偶尔 2= 有时 3 = 经常 4 = 总是

姓名	无法耐心等待	多动	注意力不集中	杂乱无章	容易分心	扰乱打扰别人
王一	0 1 2 ③ 4	0 ① 2 3 4	0 ① 2 3 4	0 1 2 ③ 4	0 1 ② 3 4	0 1 ② 3 4

7.5 The Pilot Data for the Measurement Instruments TRSC and TRSCB

In this study, prior to implementing the two instruments, I conducted a pilot test to ensure their effectiveness. The main objective of this pilot survey was to gauge how participants would respond to the survey design and determine if the questions would yield useful results. I recognised that the way respondents interpret the survey may be more important than the answers they give. Consequently, I conducted an internal pilot survey, which is a small-scale preliminary study conducted before the main research.

To carry out the pilot survey, I recruited two Year 2 teachers from each country, who completed the questionnaire without being informed that it was a pilot survey. The teachers were under the impression that the pilot questionnaire was a genuine survey, leading them to focus

primarily on the question content rather than the survey design. Upon completion of the questionnaire, I disclosed to the teachers that they had participated in a pilot study and proceeded to inquire about various aspects of the entire questionnaire. Unless any issues were identified during the internal pilot survey, the results were ultimately included in the main study, considering the respondents as initial participants. As studies always draw respondents from the target audience, an internal pilot survey provides a more realistic understanding of the type of responses researchers can anticipate when they distribute the survey to the larger group.

Pilot testing allowed the researcher to evaluate the entire questionnaire under survey conditions. The primary advantage of conducting pilot tests is the identification of problems before implementing the full survey. Specifically, pilot testing focused on examining the validity of each question, ensuring that it captures the intended information accurately. This process aligned not only with the goals and objectives of the study but also involves analysing various aspects of the entire questionnaire.

7.5.1 Accuracy of Behaviours Portrayed in Cartoon Animations

The pilot survey aimed to assess whether the behaviours portrayed in the cartoon animations accurately represent those exhibited by children with ADHD in a classroom setting. The findings indicate that the behaviours portrayed in the animations closely align with the observed behaviours of children with ADHD. Respondents recognized and identified the depicted behaviours as typical of students with ADHD, suggesting that the cartoon animations effectively captured the target behaviours.

7.5.2 Ease of Understanding the Behaviours Portrayed in Cartoon Animations

During the pilot survey, the study aimed to determine whether the behaviours depicted in the cartoon animations were easy for participants to understand and grasp. The respondents initially expressed some difficulty in comprehending and capturing the behaviours portrayed

in the first two animations. However, upon rewatching the animations, they found it much easier to understand and interpret the behaviours. The respondents attributed this initial challenge to the brevity and fast-paced nature of the cartoon animations. The quick progression of the storyline made it initially challenging for them to catch the behaviours accurately. Additionally, since the respondents were not familiar with the rating scale used in the survey, they required some additional time and practice to become familiar with it.

Once the respondents became more familiar with the rating scale and adjusted to its usage, they reported that the remaining behaviours in the animations were easier to comprehend and rate. Furthermore, when asked to describe the behaviours teachers observed in the cartoon animations, their descriptions closely matched the intended behaviours outlined in the cartoon scripts. This finding indicates that the cartoon animations effectively conveyed the intended behaviours associated with ADHD in a classroom context.

Overall, the initial difficulty experienced by the respondents in understanding and capturing the behaviours depicted in the first two animations was resolved through repeated exposure and familiarity with the rating scale. The behaviours in the animations were found to be accessible and aligned with the intended portrayal of ADHD-related behaviours. These observations confirm the effectiveness of the cartoon animations in effectively communicating the targeted behaviours associated with ADHD in a classroom setting.

7.5.3 Clarity and Comprehensibility of Questionnaire Instructions

A key objective of the pilot survey was to assess the clarity and ease of understanding of the instructions provided in the questionnaire. The data collected during this phase revealed that the instructions were generally clear and easily comprehended by the respondents. Notably, all participants successfully followed the instructions to rate the behaviours based on the

demonstrated signs. Significantly, no participants raised any questions or sought clarification on how to utilize the rating scale during the survey administration. This absence of inquiries indicates that the respondents were able to navigate the questionnaire guidelines without encountering notable confusion or ambiguity. Their ability to adhere to the instructions with minimal difficulty suggests that the instructions provided in the questionnaire were clear and easily understandable.

7.5.4 Ease of Understanding and Answering the Questions in the Questionnaire

Another crucial aspect examined in the pilot survey was the ease with which the respondents understood and answered the questions in the questionnaire. To gauge this, we solicited feedback from the participating teachers regarding both questionnaires. For the Teacher's Rating Scale of Cartoons, I requested that teachers explain the specific score they assigned to the observed behaviours in the animations. Their explanations provided valuable insights, indicating a clear understanding of the questions and rating scale. The teachers' ability to provide coherent justifications for their scores further affirmed their comprehension of the questionnaire.

Similarly, for the Teacher's Rating Scale of Children's Behaviours, I asked teachers to describe the types of behaviours corresponding to the symptoms they rated. The descriptions provided by the teachers generally aligned with the expected behaviours associated with the symptoms they rated. This alignment between the descriptions and the targeted behaviours suggests that the teachers accurately comprehended the questionnaire and its intended purpose.

The findings from the pilot survey indicate that the questions in both questionnaires were generally straightforward and comprehensible to the respondents. The participants reported minimal challenges in understanding the questions and expressed confidence in providing

relevant answers. These positive responses suggest that the questionnaire effectively elicited the desired information from the respondents, facilitating their engagement and ensuring accurate data collection.

7.5.5 Participant Perception of Rating Scales

As part of the pilot survey, I sought to understand the participants' perception of the rating scales used in the survey. To gather their feedback, we asked teachers to share their experiences and feelings about using these rating scales. Here are the key findings.

Semantic Differential Scale (SDS): Teachers reflected positively on the use of the Semantic Differential Scale. They found it to be straightforward and easy to use. The graphical representation of the scale, presented as a line, aided their understanding and visualization of their position on the scale. Teachers mentioned that they only needed to remember which end represented a low rating and which end represented a high rating. By relying on their feelings, rather than struggling with numerical values, they were able to rate the behaviours in the cartoons effectively. The user-friendly nature of the scale facilitated their engagement with the survey.

Likert Scale for Rating Children's Behaviours Frequency: Teachers provided valuable feedback regarding the Likert scale used to rate the frequency of children's behaviours. One teacher mentioned that she did not use the score "0=Never" because, in her experience, the behaviours in the questionnaire were common behaviours among children, and almost all the children in her class exhibited them. Another teacher indicated that she faced some confusion in distinguishing the exact differences between "2=Often," "3=Very Often," and "4=Almost Always." However, she mentioned that she had no difficulty in rating the children's behaviours since she associated the related numbers as scores for rating.

Overall, the feedback from teachers indicated that the majority did not encounter significant confusion or difficulties in using the rating scales. Participants found the scales to be clear and appropriate for accurately expressing their opinions or perceptions. This feedback suggests that the rating scales employed in the survey effectively captured the desired feedback from the participants, contributing to the robustness of the data collected. Furthermore, the positive reception of the rating scales underscores their suitability for assessing the various aspects under investigation. The seamless integration of these scales into the research process enhances the reliability and validity of the study's findings. As a result, the feedback provided by teachers reinforces the confidence in the utility of the chosen methodology and serves to validate the overall research endeavour.

7.5.6 Participants' Feedback on Questionnaire Length

I sought participants' feedback on the length of the questionnaire. The responses provided a mixed perception regarding the questionnaire length.The majority of participants (three teachers) considered the questionnaire to be of an appropriate length, indicating that they did not encounter any significant issues or exhaustion while completing it. However, one teacher expressed her opinion that rating 10 children's behaviours at once was intense and exhausting, potentially leading to inaccuracies in her ratings. She suggested that a maximum of seven children would be more manageable and allow for more accurate ratings.

During the pilot study, I also recorded the time each teacher spent on completing the questionnaires. The recorded times ranged from 26 to 35 minutes. As the survey was designed to be completed within half an hour, the recorded completion times generally aligned with the intended timeframe. Overall, while the majority of participants found the questionnaire length to be acceptable, one teacher highlighted the potential challenges associated with rating a larger

number of behaviours. Taking into account the feedback and the recorded completion times, the questionnaire design appears to align with the intended length.

In conclusion, the pilot survey findings demonstrate that the cartoon animations effectively portrayed behaviours associated with ADHD in a classroom setting. The questionnaire instructions were clear and easily understood, facilitating participants' comprehension and response accuracy. The questions in the questionnaire were generally comprehensible and straightforward, with the rating scales not causing significant confusion. While the questionnaire length received mixed feedback, participants acknowledged the relevance of the questions. These findings provide valuable insights for refining the survey design and informing the main study, ensuring a robust and comprehensive assessment of behaviours related to ADHD in children.

7.6 Conclusion

The journey of developing measurement instruments to assess teachers' leniency towards ADHD symptoms and children's ADHD behaviours has been a multifaceted and intricate process. In this chapter, we embarked on exploring the various stages of this development, commencing with the creation of cartoon animations to accurately portray children's ADHD behaviours in a culturally unbiased manner. This comprehensive approach involved conducting classroom observations, engaging in teacher interviews, and conducting a thorough review of rating scales, all contributing to the development of an extensive ADHD behavioural bank with rich and detailed contexts. Building upon this abundant material, the study collaborated with professional animators to craft a captivating series of cartoon animations that brought the children’s ADHD behaviours to life in an authentic classroom setting.

Subsequently, the chapter focus shifted towards the development of two essential measurement instruments: the Teacher's Rating Scale of Cartoons questionnaire (TRSC) and the Teacher's Rating Scale of Children's Behaviours (TRSCB). The TRSC was meticulously designed to gauge teachers' leniency towards six specific ADHD symptoms: Can't wait, Fidgety, Inattentive, Disorganized, Distracted, and Disruptive. On the other hand, the TRSCB was developed to evaluate children's ADHD behaviours in relation to these six symptoms, offering valuable insights into how they exhibit these behaviours in both countries. To validate and fine-tune the measurement instruments, an internal pilot survey was conducted to gain a realistic understanding of the anticipated responses from a larger group of participants. The insights gained from the pilot data further refined the measurement instruments, ensuring their accuracy, reliability, and sensitivity in capturing the intended information.

In this chapter, it was clear that we were fully committed to creating reliable and culturally unbiased measurement instruments, as demonstrated by the thorough and careful development process. These instruments hold the potential to serve as indispensable tools for investigating rater bias and facilitating cross-cultural comparisons in ADHD assessments among teachers from China and England. In the following chapter, I will delve into the intricate details of our research methods and methodologies, providing a comprehensive and insightful exploration of this study.

Chapter 8 Research Methods

8.1 Introduction

The methods and methodology chapter serves a vital purpose in the research process as it outlines the strategies and approaches employed to address the research questions and achieve the objectives of the study. This chapter plays a significant role in ensuring the reliability, validity, and integrity of the research findings. In this chapter, I present the methodological framework that will guide the investigation into the research questions identified in previous chapters. I will provide a comprehensive overview of the research design, sampling method, data collection, measurements and analysis method involved in this study.

8.2 Research Design

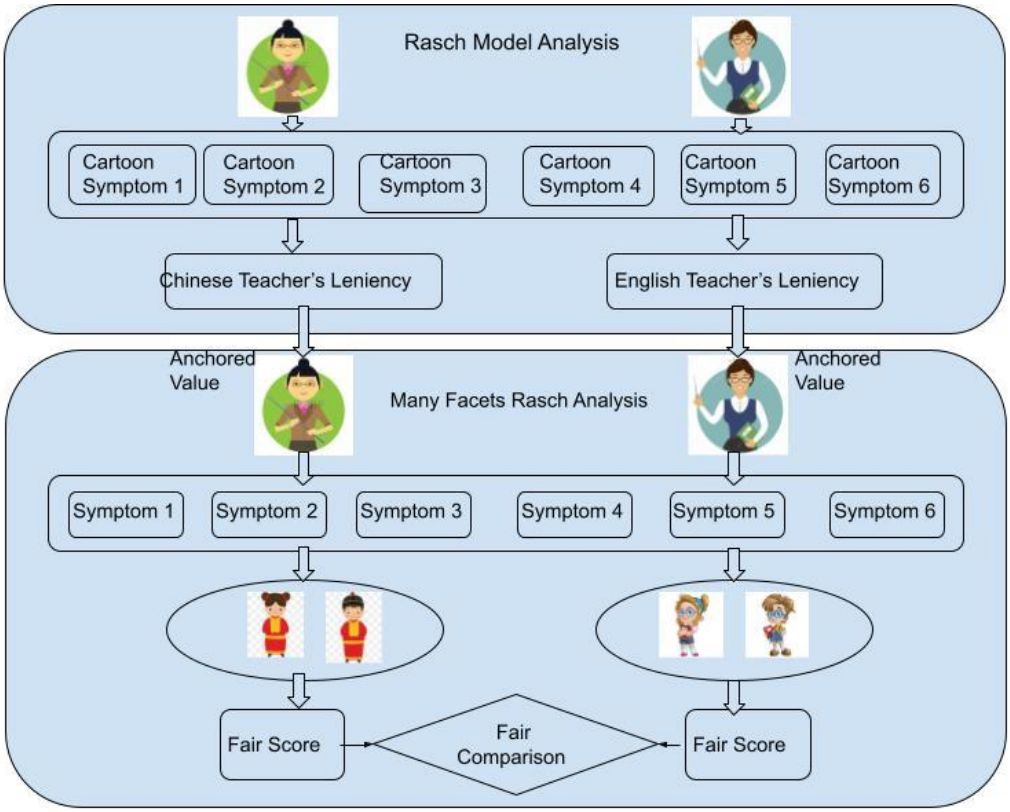
In previous chapters, the study explained that judgments about children's inattentive, hyperactive, and impulsive behaviours can vary based on cultural and social contexts. When comparing ADHD symptoms between children from different cultural backgrounds, it becomes challenging to determine if the differences are due to the children themselves or the raters' bias. To ensure a fair comparison, it needs to measure the leniency of Chinese and English teachers in rating children's ADHD behaviours. This measurement will help us account for potential differences in leniency and provide more accurate prevalence rates of ADHD in children across different cultures and societies. Therefore, the first question is how to measure the leniency of Chinese and English teachers in rating children's ADHD behaviours?

There are ways to identify rater bias in the rating of children's ADHD behaviours across different cultures using a standardised rating target such as written vignettes, videotape vignettes or objective observations. However, they have limitations for this study, as discussed before. Therefore, this study adopted cartoon animations as media to develop six episodes of cartoon animations to represent children's behaviours of six ADHD symptoms (Can't wait,

Fidgety, Inattentive, Distractive, Disorganised and Disruptive) as standardised vignettes for Chinese and English teachers to observe and rate. The teachers filled out a questionnaire (Teacher's Rating Scale of Cartoon) about their tolerance towards the behaviours they observed in the cartoon animations. With the cartoon animations and the rating scale, the study could measure teachers' leniency in rating children's ADHD behaviours across the two countries. With Rasch model analysis, the study could produce a unidimensional equal-interval scale to measure teachers' leniency towards children's ADHD behaviours in two countries.

For investigating how to make appropriate adjustments to ratings of children's ADHD behaviours to improve fairness in cross-cultural comparison? The study developed the Teacher's Rating Scale of Children's Behaviour to measure children's ADHD behaviours across China and England. Teachers from both countries were asked to rate 10 children randomly selected from their class with the rating scale. Furthermore, the study has adopted Many-facet Rasch Measurement (MFRM) (Linacre, 1989) to produce fair comparisons by introducing teachers' leniency as anchored facet in the analysis. The research design is shown in *Figure 17.*

Figure 17 The research design



8.3 Measurements

In the study, I utilised two instruments to gather data. The first instrument is the Teachers' Rating Scale of Cartoons (TRSC). The second instrument is the Teachers' Rating Scale of Children's Behaviours (TRSCB). This section will delve into various aspects of the two instruments, including instrument description, purpose, scoring and interpretation and reliability.

8.3.1 The Teachers' Rating Scale of Cartoons (TRSC)

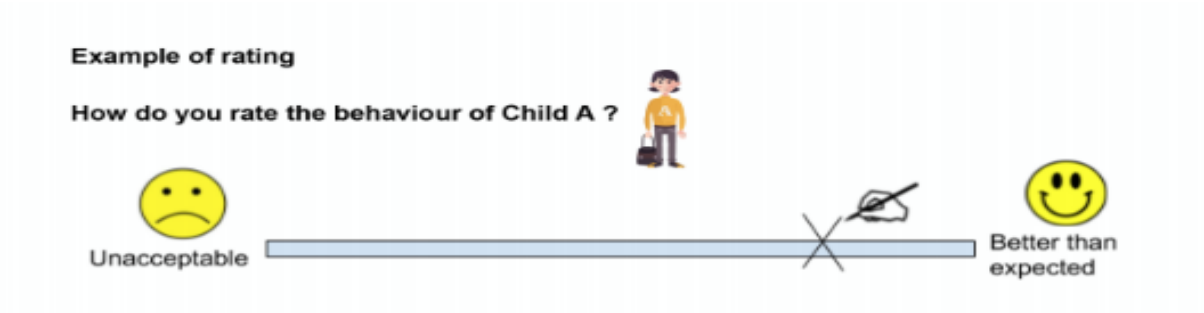
Description

The first part involves a series of six episodes of cartoon animations. These animations specifically depict the behaviours associated with six ADHD symptoms: "Can't wait," "Fidgety," "Inattentive," "Distractive," "Disorganised," and "Disruptive." The cartoons are designed to portray these behaviours in a classroom setting, providing a visual representation for the participating teachers to observe. The example of cartoon animations is in *Figure 18*.

Figure 18 Example of cartoon animations

The second part of the TRSC is a questionnaire that incorporates semantic differential scales (SDS). These scales are presented as bipolar lines, with the negative attitude positioned on the left end and the positive attitude on the right end. The participating teachers are asked to indicate their attitude toward each behaviour they observe in the cartoon animations by placing a cross sign at any point along the line. This allows them to express their attitude or perception of the behaviour, ranging from negative to positive. The example of TRSC questionnaire is in Figure *19*.

Figure 19 TRSC scale



The study developed six cartoons to present six different ADHD symptoms. Each of them has five cartoon characters to present five different behaviours of each symptom. The symptoms

and behaviours are listed in *Table 21*. In the survey, I asked teachers to rate the behaviours of the cartoon characters on a semantic differential scale (length: 10cm) by putting a cross on the scale depending on their acceptance of the behaviours they observed (see *Figure 19*). One end of the scale is "Unacceptable", and the other end is "Better than expected". Teachers can put crosses at any position on the scale. If they feel negative about the behaviour, they might put the cross close to the "Unacceptable" end. If they feel positive about the behaviour, they might put the cross close to the "Better than expected" end. They might put the cross in the middle of the scale if they feel neutral about the behaviour. When extracted data from the surveys, I measured the distance between the cross and the end of "Unacceptable". The longer the distance is, the more accepting the teacher is of the behaviour.

Table 21The animated symptoms and related behaviours list

Cartoon/Symptom	Behaviours
Can't Wait	A1 Leaving seat and asking for sweets
	B1 Shouting out for sweets
	C1 Waiting for
	D1 Leaving seat and eating sweets
	E1 Leaving seat and touching
Fidgety	A2 Moving body on seat
	B2 Fiddling with a pencil and tapping feet
	C2 Moving chair
	D2 Sitting still
	E2 Fiddling with other's hair
Inattentive	A3 Doodling
	B3 Tearing paper
	C3 Daydreaming

	D3 Eating snack
	E3 Concentrating
Disorganised	A4 Late for school
	B4 Messy table
	C4 Tidy table
	D4 Looking for things
	E4 Nothing on the table
Distracted	A5 Having a look and back to the lesson
	B5 Looking at the butterfly all the time
	C5 Playing with the butterfly but staying on seat
	D5 Shouting out
	E5 Leaving seat to chaise the butterfly
Disruptive	A6 Concentrating on writing
	B6 Taking other's pencil without asking
	C6 Tickling another child
	D6 Doodling on other's writing
	E6 Disturbing the teacher

Purpose

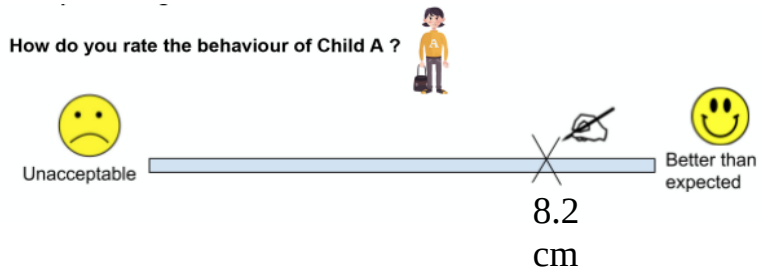
The purpose of this instrument is to gauge teachers' leniency towards children's behaviours associated with the six ADHD symptoms. It is essential to measure teachers' leniency for cross-cultural comparison studies to account for potential cultural biases or variations in their perceptions and judgments of children's behaviours. To facilitate data collection, the TRSC was administered as the initial instrument during the survey. This approach ensured that the teachers have a general impression of what type of behaviours each symptoms associated

before they rated children in their own class. This will facility the second instrument (TRSCB) administration.

Scoring and Interpretation

In the scale, there are 30 animated children's ADHD-type behaviours in six cartoon animations. English and Chinese teachers were asked to watch the cartoons and rate these animated behaviours according to their acceptance. The rating scale is a semantic differential scale shown in *Figure 20* **.**

Figure 20 TRSC rating scale example



Their ratings were extracted from the questionnaires by measuring the location of teachers' marks on the semantic differential scale; for example, 8.2 cm equals 8.2 on the measurement scale. Therefore, the original readings from the measurement are in decimal format. However, the Winsteps software only analyses ordinal data, not decimal data. Before conducting Rasch analysis using the Winsteps software, the questionnaire data were converted into ordinal data (0, 1, 2). The data conversion scheme is detailed in *Table 22*. The investigation into the reliability and function of the Teachers' Rating Scale of Cartoon was carried out separately for the Chinese and English samples.

Table 22 The data converted scheme

Survey Data (Decimal)	Winsteps Data (ordinal)
[0, 3.3]	0
[3.4, 6.7]	1
[6.8, 10]	2

All the data from the questionnaires were measured and verified to ensure the accuracy of the information extracted. Additionally, during the process of transferring this data to other digital formats (such as Excel, Txt, and SPSS), the input data were meticulously reviewed to confirm the accuracy of the conversion. Furthermore, when converting data from a continuous scale into a categorical scale for analysis using Winsteps and Facets, all the transformed data underwent a thorough double-check to ensure the accuracy of the transformation. An identification and labelling system for teachers, children, cartoon behaviours, and children's behaviours was developed, as outlined in *Table 23.*

Table 23 ID and labelling system

	Teacher	Children
China	t001-t109	cc001-cc600
England	t01-t21	ce001-ce240
Cartoon	Behaviours	
Can’t Wait	A1, B1, C1, D1, E1	
Fidgety	A2, B2, C2, D2, E2	
Inattentive	A3, B3, C3, D3, E3	
Disorganised	A4, B4, C4, D4, E4	
Distracted	A5, B5, C5, D5, E5	
Disruptive	A6, B6, C6, D6, E6	

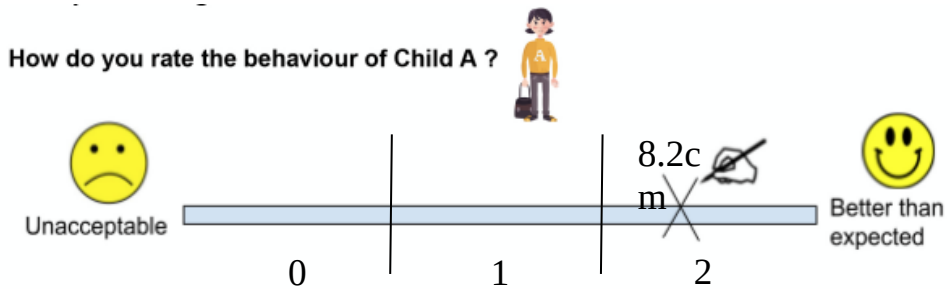
8.3.2 Reliability of the Teacher's Rating Scale of Cartoons (TRSC)

The Reliability and function Investigation of Chinese Sample

There were 30 items on the scale, and this scale measured 106 Chinese teachers. The reliability and function investigation were carried out with Rasch analysis. It includes investigation of category function, summary statistics, Wright map, misfit person, misfit item, dimensionality and local dependence.

Category Function. As Winsteps analyses ordinal data, the decimal data extracted from the semantic differential scale had to be converted into ordinal data, as illustrated in *Figure 21*. To determine whether the categorical scale (0, 1, 2) I have derived from the decimal data is functional for measuring the Chinese sample, I need to examine the category function.

Figure 21 Example of convert rating of decimal data (8.2) to ordinal data (2)



In the analysis of the category function for this rating scale, several key indicators were considered. According to established guidelines (Linacre & Wright, 1994), INFIT and OUTFIT MNSQ values falling within the range of 0.6 to 1.4 are indicative of productive performance for rating scale measurement. Examination of *Table 24* reveals that the INFIT MNSQ values span from 0.94 to 1.04, while the OUTFIT MNSQ values range from 0.97 to 1.13. These results suggest that the rating scale is effectively capturing the characteristics of the sample being measured.

Furthermore, Linacre (1999) recommended a minimum threshold of 1.4 logit for the Andrich threshold, which signifies the minimal advancement in step difficulties between adjacent categories. In *Table 24*, the Andrich Threshold between category '1' and category '2' is calculated to be 1.22 logit (.61-(-.61) = 1.22). While this value indicates a relatively small gap between these two categories, suggesting that the step difficulties are slightly lower than the recommended threshold, further analysis is required.

Figure 22 presents probability curves that shed light on the categorical scale's performance in measuring the sample. These curves provide an insightful summary of how respondents utilised the response categories (0, 1, 2) in their assessments. The horizontal x-axis signifies the item's level of difficulty, while the vertical y-axis represents the expected likelihood of endorsement for each category when teachers assigned ratings to items. To be considered as functioning reasonably well, it is expected that the highest probability peak for any category should have a minimum threshold value of 0.5.

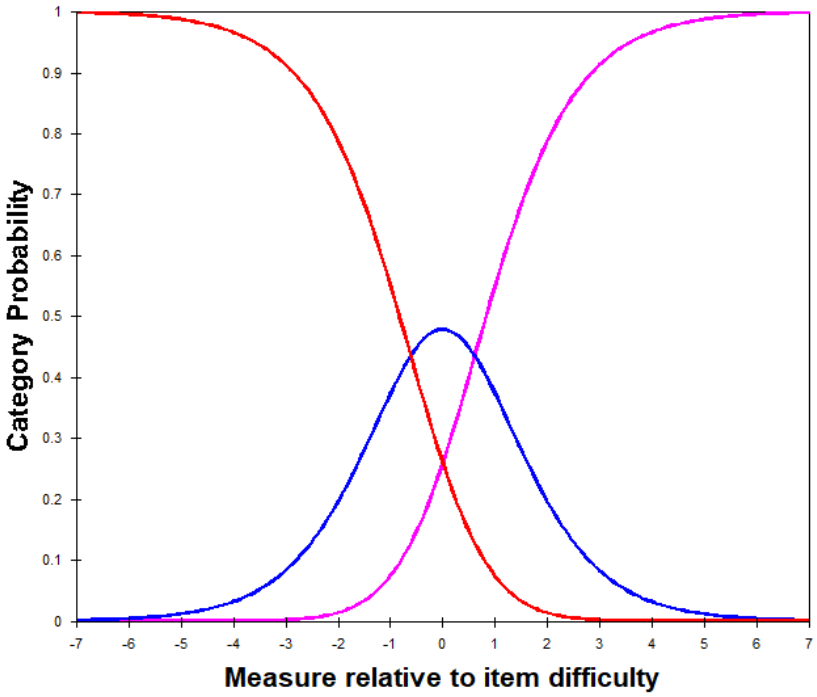
In *Figure 22*, three distinct peaks represent the probabilities of endorsement for the three categories. The red line corresponds to category 0, the blue line to category 1, and the purple line to category 2. Overall, these probability curves demonstrate that the categorical scale functions reasonably well for measuring the sample, even though the step difficulties between adjacent categories may be slightly lower than the recommended threshold. This suggests that the categories within the rating scale effectively discriminate between different levels of the latent trait, making the scale a suitable tool for measurement in the Chinese sample.

Table 24 Summary of category function

Category Label	INFIT MNSQ	OUTFIT MNSQ	Andrich Threshold	Category measures
0	.97	1.13	None	-1.89

1	.94	.97	-.61	.00
2	1.04	1.10	.61	1.89

Figure 22 Category probabilities



Summary Statistics of Persons and Items. In Rasch measurement, person ability and item difficulty are calibrated on the same interval-level measurement scale; they are conjoined. Therefore, ability and difficulty can be compared directly on one scale. The summary statistics of persons and items give us a general view of how the items perform together to construct a measurement to measure persons' abilities. Software Winsteps reports person and item reliability and separation with different implications. Reliability implies the "reproducibility of relative measure location" (Bond & Fox, 2015). High reliability indicates a high possibility that the person (item) with a high estimated measure does have a higher ability (difficulty) than the person (item) with a low estimated measure.

Person separation indicates how sensitive the instrument is to classify people with different levels of abilities. Low person separation (<2, person reliability <.8) indicates that the instrument is not sensitive to identify high ability and low ability persons (Bond & Fox, 2015). The instrument might need more or better items to improve. Item separation indicates the hierarchy of item difficulties. Low item separation (<3, item reliability<.9) implies that person's sample is not big enough to give enough information to estimate item difficulties (Bond & Fox, 2015). Increasing the sample size with a wide range of people's abilities will help increase the item separation.

The reliability is calculated with *Equation 4*. In Winsteps the reliability and separation indexes were reported with a lower bound value (REAL RMSE) and an upper bound value (MODEL RMSE). In my study, I only report the lower bound value REAL RMSE.

Equation 4

$$\text{Reliability}=(\text{OV}-\text{EV})/\text{EV}$$

OV = observed variance of item difficulty /person ability measures

EV = mean of squared standard errors of item difficulty/person ability measures

In *Table 25,* the average person's measure (Mean) is -.35 logit. Although the person's measure ranges from 1.59 logit to -3.13 logit, the standard deviation of 0.86 logit for the person's estimates indicates that most of the population spread in a relatively narrow band of 1.72 logit. The reliability of the teachers' attitude estimates (Person Reliability) is .81, and the Separation (Real RMSE) is 2.08, which is acceptable. It indicates that the scale discriminates between high-ability and low-ability people as it was designed.

In *Table 25*, the average (Mean) measure of items is located at 0 logit as default. The difficulty of items is in a wide range from -3.75 logit to 2.52 logit. Most item difficulties are located in a standard deviation of 1.74 logit. It indicates that items spread in a wide range of difficulties to measure the trait. Item reliability and separation indicate the ability of the measurement to define a distinct hierarchy of items along the measured trait. The higher the number, the more confident I can apply the measurement across other samples and get the same order of item estimates. The item reliability (Item Reliability) is .98, and the separation is 8.06. It suggests that I are confident that I can rely on this order of item estimates for other suitable samples.

Table 25 Summary statistics of persons and items

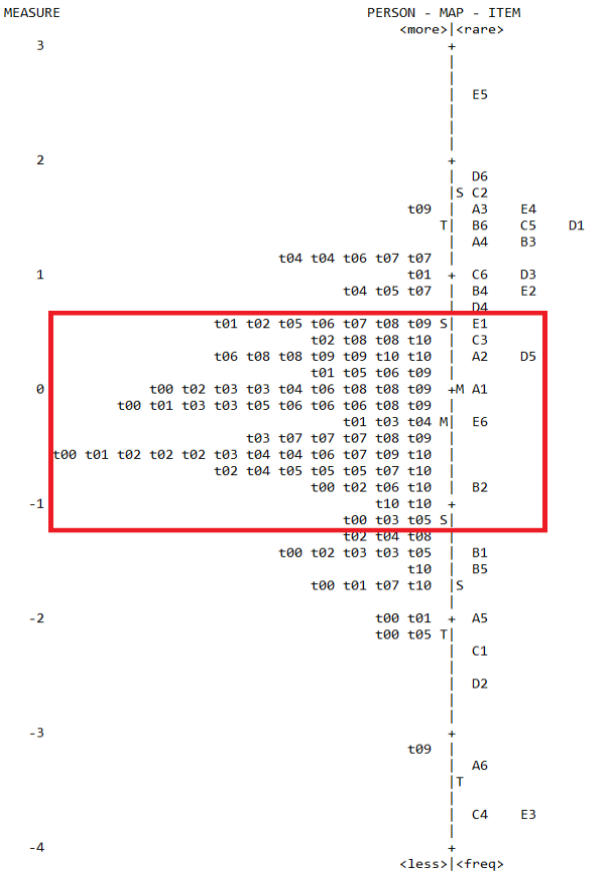
	Mean	P.SD	Min	Max	Separation	Person/Item Reliability
Persons	-.35	.86	-3.13	1.59	2.08	.81
Items	.00	1.74	-3.75	2.52	8.06	.98

In general, for the Chinese sample, the summary statistics of persons show that the reliability of persons' estimates is not high but good enough to discriminate different levels of abilities. It might be due to limited items (30 items) for a big sample (106 persons). The identification of different levels of persons' abilities is limited. The summary statistics of items show that the reliability of the items' estimates is very high.

Wright Map. The Wright Map of *Figure 23* shows us the person and item relations visually on one logit scale. The estimates of persons' abilities and items' difficulties are displayed on the side of the logit scale. Items are indicated by the item number (for example, A1), and a person's estimate is represented by the person's number (t01). This map is also called an Item-Person Map. The Wright map shows that 30 items spread from bottom to top of the scale and covered all the abilities of the sample, and the spreading of persons' abilities almost matched the items.

It indicates that the measurement is not too easy or too difficult. Most of the persons' estimates are located between two 'S' in the red square on the left side of the scale. However, only 7 items in this area distinguish most of the persons' abilities. That is one of the reasons why the reliability of persons' estimates is lower than the reliability of item estimates. In other words, the performances of 106 persons tell us more reliable information about items than 30 items tell us about the 106 persons. More items would improve the discrimination of persons' abilities.

Figure 23 Wright map of the Chinese sample



Summary of Items and Persons Misfit statistics of Chinese Sample.

Five indexes in the item and person misfit table show whether the items fit the model. INFIT is inlier-pattern-sensitive fit statistics, which is more sensitive to unexpected behaviour affecting responses to items near the person's measure level. OUTFIT is outlier-pattern-sensitive fit statistics, more sensitive to people's unexpected behaviour on items far from the person's measure level. MNSQ is a mean-square statistic that indicates the size of a

misfit. *Table 26* interprets the MNSQ statistics, and *Table 27* shows the reasonable Item Mean-square Ranges for INFIT and OUTFIT. INFIT and OUTFIT MNSQ values that range from .6 to 1.4 are considered acceptable for the rating scale. (Bond & Fox, 2015. P273).

ZSTD statistic corresponding to each mean-square shows the probability associated with the null hypothesis: "These data fit the Rasch model". If ZSTD is more extreme than ±1.96, then there is "statistical significance", and the null hypothesis is rejected. Therefore, INFIT and OUTFIT ZSTD values within ±2 are acceptable. The PTMEASURE-AL CORR is a point-measure correlation. It investigates whether responses to the items align with the abilities of persons. Therefore, negative correlations usually indicate that the responses to the item contradict the direction of the latent variable. These items with negative correlations are problematic.

Table 26 Interpretation of parameter-level mean-square fit statistics

>2.0	Distorts or degrades the measurement system
1.5 - 2.0	Unproductive for construction of measurement, but not degrading
0.5 - 1.5	Productive for measurement
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

Note. (Bond & Fox, 2015. P272) Bond, T., & Fox, C. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.)

Table 27 Reasonable items mean-square ranges for INFIT and OUTFIT

Type of Test	Range
MCQ (High stakes)	0.8 - 1.2
MCQ (Run of the mill)	0.7 - 1.3
Rating Scale (survey)	0.6 - 1.4
Clinical observation	0.5 - 1.7
Judge (agreement encouraged)	0.4 - 1.2

Note. (Bond & Fox, 2015. P273) Bond, T., & Fox, C. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.)

In *Table 28*, INFIT MNSQ values range from 0.67 to 1.51, which is very close to the acceptable range, with only three items (D2, C1, C4) slightly out of the range. OUTFIT MNSQ range from 0.67 to 2.01, with four items (D2, C1, C4, E6) out of the range. The INFIT ZSTD range from -3.22 to 2.32, and the OUTFIT ZSTD range from -2.72 to 3.09. When deciding which items are more problematic, INFIT MNSQ out of acceptable range and ZSTD values greater than 2.0 should be first investigated because these underfitting items impact item and person estimates more than items with high OUTFIT ZSTD. In *Table 29* there are only three items (D2, C1, E6) with INFIT ZSTD greater than 2.0, and both INFIT and OUTFIT MNSQ are out of the acceptable range. However, there are no negative values of PTMEASURE-AL CORR, which indicates that all the items align with the abilities of persons.

Table 28 Item misfit statistics

Items	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
	.67 - 1.51	-3.22 - 2.32	.67 - 2.01	-2.72 - 3.09	.11 - .61

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

Table 29 The misfit items statistics

Items	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
D2	1.51	2.32	2.01	3.09	.11
C1	1.41	2.13	1.88	3.13	.14
E6	1.34	2.85	1.68	4.85	.34

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

The misfit items' Item Characteristic Curve (ICC) are shown in *Figure 24*, *Figure 25* and *Figure 26*. ICC shows how the probability of getting an item right change with the underlying trait. The red line is the mode line which is the same for all the items. The blue line is the observed line or empirical line. The two green lines along both sides of the red line are the boundary lines that represent the upper and lower 95% of two-sided confidence intervals. When the blue line goes out of the boundaries, it indicates that some unmodeled source of variance may be present in the observations.

The item E6 (disturbing the teacher) curve looks different from the others. The curve shows that this item may have two dimensions. It indicates that some teachers think this behaviour is good while others think it is unacceptable. From the unexpected responses table, I found that a very accepting teacher (t051) rated this behaviour unacceptable, while some less acceptable teachers (t094, t06, t072, t013) rated this behaviour acceptable. During the survey, some teachers expressed that they would like their students to interact with them during lessons, including disturbing their lectures when students have questions. While other teachers expressed that they do not like being disturbed by students while they are talking. I think that teachers' rating about whether this behaviour is acceptable or not depends on teachers' teaching style. Therefore, I kept item E6 for further analysis.

Figure 24 ICC curve of item D2 (sitting still)

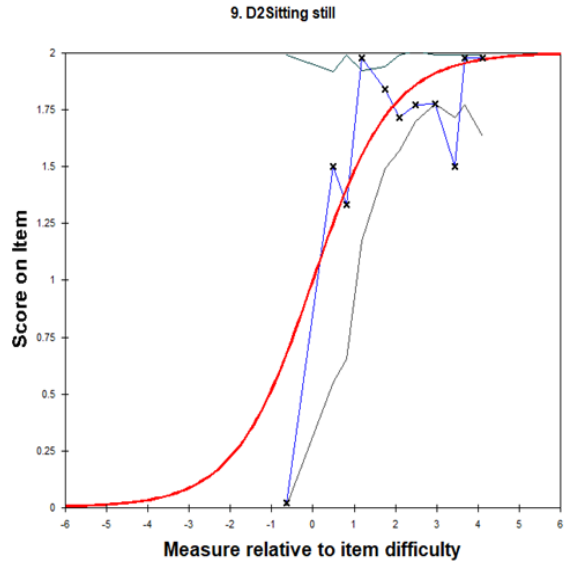


Figure 25 ICC curve of item C1(waiting for)

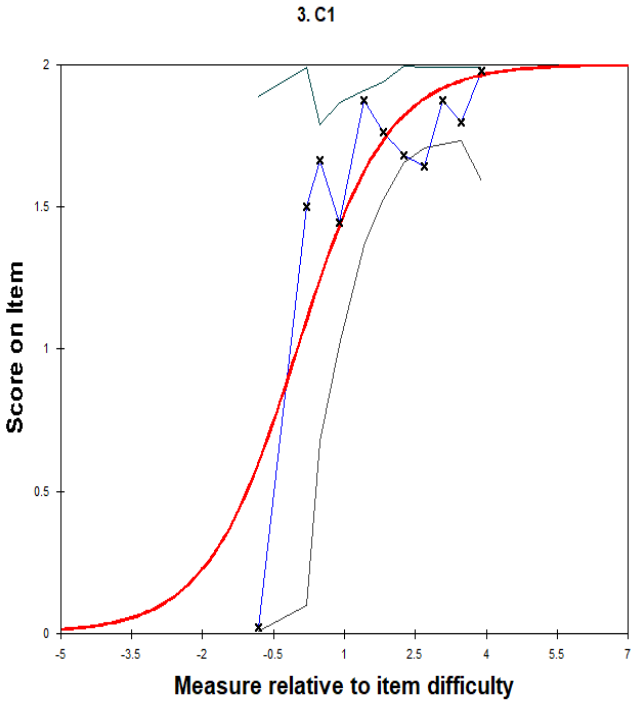
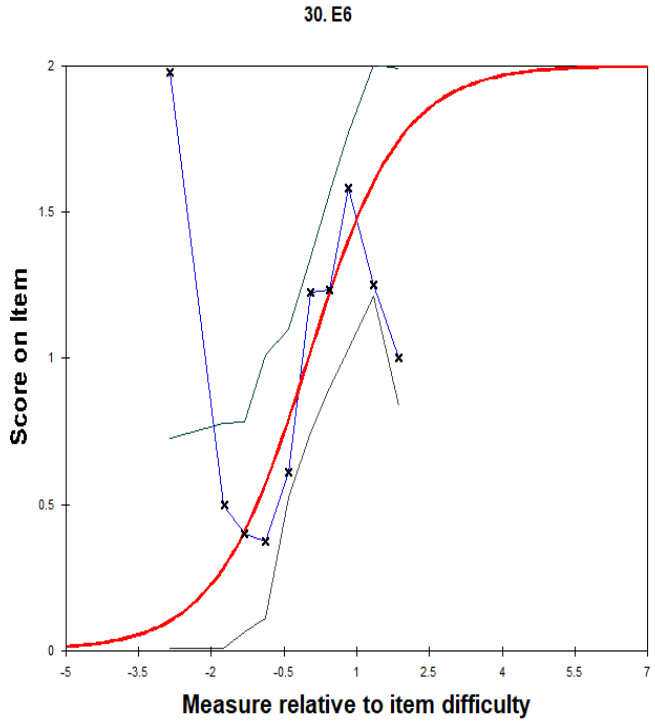


Figure 26 ICC curve of item E6 (Disturbing the teacher)



In *Table 30*, the person's INFIT MNSQ values range from .47 to 2.31, and INFIT ZSTD values range from -2.62 to 3.37. OUTFIT MINSQ values range from .45 to 4.84, and OUTFIT ZSTD values range from -1.69 to 5.03. Although there are many out-of-range values, the INFIT MNSQ values out of acceptable range and INFIT ZSTD values greater than 2 are the most concerned ones. There are 14 persons with INFIT MNSQ values out of the acceptable range and INFIT ZSTD values greater than 2. There are 8 persons with OUTFIT MNSQ values out of the acceptable range, and OUTFIT ZSTD values are greater than 2. Therefore, 22 persons are outliers with either INFIT, OUTFIT, or both out of range.

Table 30 Person misfit statistics

Person	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
	.47 - 2.31	-2.62 - 3.37	.45 - 4.84	-1.69 - 5.03	.06 - .90

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

Dimensionality and Local Dependence. Residuals are the differences between the observed and expected scores generated from the Rasch model (Bond & Fox, 2015). If a test is unidimensional, the residuals should be distributed randomly, and no pattern should exist. If there is a pattern, the test has a meaningful sub-dimension and is not unidimensional. Rasch PCA of residuals is looking for any meaningful sub-dimension in the residuals. *Table 31* Standardized Residual Variance shows that the Rasch dimension explains 59.0% of the variance in data which is good. However, the unexplained variance in the first contrast is 3.63 units of Eigenvalue. The first contrast with an eigenvalue greater than 2 should be a concern (Linacre, 2017). One Eigenvalue unit equals 1 item, which means that in our scale, there are almost 4 items the residuals show a pattern of potential sub-dimension.

Figure 27 Standardised residual first contrast plot shows the loading of the first contrast of items. The X-axis represents the difficulty measures of items, and Y-axis represents the loadings on the conceptual component in the unexplained variance. The positive loading items show a cluster of the largest correlation with the latent sub-dimension (Linacre, 2017). In *Figure 27***,** the highlighted items A, B, C, and D have high loadings of over .5. *Table 32* shows the details of the items with the largest loading in the first contrast. Examining the content of

these four items shows that item A (E3) and item D (D2) have similar behaviour: focusing. Item B (A5) and Item C (B5) are from the same scenario, which shows how children behave when a butterfly flies into the classroom. The behavioural difference between A5 and B5 is relatively subtle, and behaviour A5 shows more focus than B5. These four items generally represent children's focusing behaviour on different scenarios. Meanwhile, items A, B, C, and D are more acceptable behaviours for Chinese teachers than other behaviours as the item difficulties are negative on the plot. They do not form a different dimension of the latent the instrument tries to measure.

Table 33 shows the items' largest standardized residual correlations to identify any dependent items. The paired items with high positive correlations indicate local dependency. As "common variance= r2", the paired items A5 and B5 only have 37% of the variance in their residuals in common. Linacre (2017) recommended that correlations be around 0.7 before we are concerned about dependency. Therefore, there is no evidence of dependent items.

Table 31 Standardized Residual Variance

	Eigenvalue	Observed	Expected
Total raw variance in observations	73.20	100%	100%
Raw variance explained by measures	43.20	59.0%	58.6%
Raw variance explained by persons	11.30	15.3%	15.2%
Raw variance explained by items	31.99	43.7%	43.4%
Raw variance unexplained variance (total)	30.00	41.0%	41.4%
Unexplained variance in 1st contrast	3.63	5.0%	12.1%

Figure 27 Standardized residual first contrast plot

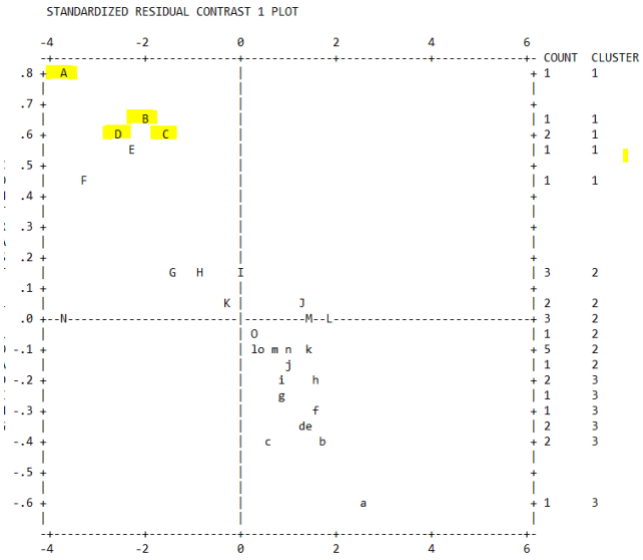


Table 32 Content of 1st contrast items with largest loadings

Label in plot	Label in scale	Loading	Content
A	E3	.78	Concentrating
B	A5	.67	Having a look and back to lesson
C	B5	.59	Looking at the butterfly all the time
D	D2	.58	Sitting still

Table 33 Largest standardized residual correlations used to identify dependent item

Correlation	Item	Item
.61	A5 Having a look and back to lesson	B5 Looking at the butterfly all the time
.57	D1 Leaving seat and eating sweets	E1 Leaving seat and touching sweets
.46	D2 Sitting still	E3 Concentrating
.40	E3 Concentrating	A5 Having a look and back to lesson
.39	C1 Waiting for	E3 Concentrating
.37	C1 Waiting for	D2 Sitting still

In summary, the Rasch Model analysis of the rating scale for the Chinese sample yielded valuable insights. The category function analysis, based on INFIT and OUTFIT MNSQ values, revealed that the scale effectively measured the sample, falling within the recommended range for productive performance. While the Andrich Threshold indicated a slight deviation, probability curves demonstrated the scale's reasonable functioning with the three-point category. Person and item reliability and separation statistics showcased the scale's ability to

discriminate between different abilities, with some room for improvement in item discrimination. The Wright Map illustrated a balanced distribution of items and persons along the logit scale. However, additional items could enhance discrimination. Item and person misfit analysis identified outliers needing further investigation. Dimensionality analysis suggested a potential sub-dimension related to focusing behaviour but not representing a distinct latent trait. Importantly, there was no evidence of dependent items. Overall, the rating scale offers a reliable and functional tool for measuring Chinese teachers' perceptions, with some minor areas for refinement.

The Reliability and Function Investigation of England Sample

23 English teachers were measured by the instrument.

Category Function. In order to know whether the categorical scale (0, 1, 2) for measuring the English sample is productive, it needed to investigate the category function. INFIT and OUTFIT MNSQ from .6 to 1.4 are productive for rating scale measurement (Linacre & Wright, 1994). In *Table 34*, the INFIT MNSQ values range from .93 to 1.15 and OUTFIT MNSQ values range from .86 to 1.14. It indicates that the scale is productive for measuring the sample. Linacre (1999) recommended a minimum of 1.4 logit for the Andrich threshold. In *Table 34*, the ANDRICH THRESHOLD between category '1' to category '2' is 1.64 logit (.82-(-.82) =1.64). It means that the necessary minimum of advance in step difficulties between these two categories is satisfied.

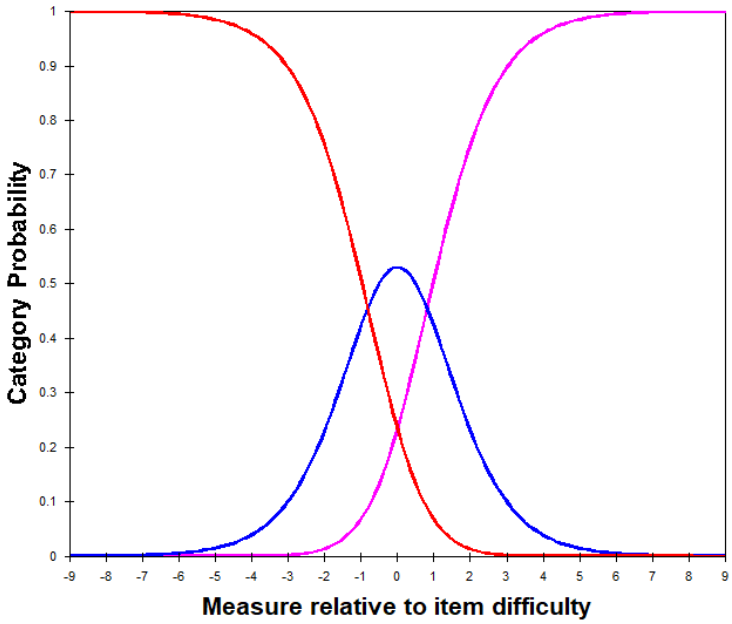
In *Figure 28,* the probability curves show that the categorical scale functions reasonably well for measuring the sample. The curves summarise how the sample uses the response (0,1,2) in the measurement. The horizontal x-axis represents the item's difficulties. The vertical y-axis represents the expected possibility of endorsement of any given category when a teacher gives a rating to an item. The minimum threshold value for the highest probability peak for any

category is expected to be.5 for the category to be interpreted as functioning reasonably well. In *Figure 28,* three peaks represent the probabilities of endorsement of three categories. In general, the category is functioning well in the English sample.

Table 34 Summary of category function of English sample

Category Label	INFIT MNSQ	OUTFIT MNSQ	Andrich Threshold	Category measures
0	.93	.97	None	-2.04
1	.94	.86	-.82	.00
2	1.15	1.14	.82	2.04

Figure 28 Category probabilities



Summary Statistics of Persons and Items. In *Table 35*, the average person's measure (Mean) is -.38 logit. Person's measure is in a wide range of 1.75 logit to -3.63 logit with a standard deviation (P.SD) of 1.20 logit for person's estimates, indicating that most of the population spread in a relatively good range of 2.40 logit on the scale. The reliability of the teachers' attitude estimates (Person Reliability) is .87, and Separation is 2.57, which is good.

As shown in *Table 35*, the average (Mean) measure of items is located at 0 logit as default. The difficulty of items is in a wide range from -5.13 logit to 4.09 logit. Most item difficulties are located in a standard deviation of 2.77 logit. It indicates that items spread in a wide range of difficulties to measure the trait. Item reliability and separation indicate the ability of the measurement to define a distinct hierarchy of items along the measured trait. The higher the number, the more confident I can apply the measurement across other samples and get the same order of item estimates. The item reliability is .95, and the separation is 4.33. They both are high. It suggests that the study can rely on this order of item estimates for other suitable samples.

Table 35 Summary statistics of persons and items

	Mean	P.SD	Min	Max	Separation	Reliability
Persons	-.38	1.20	-3.63	1.75	2.57	.87
Items	.00	2.77	-5.13	4.09	4.33	.95

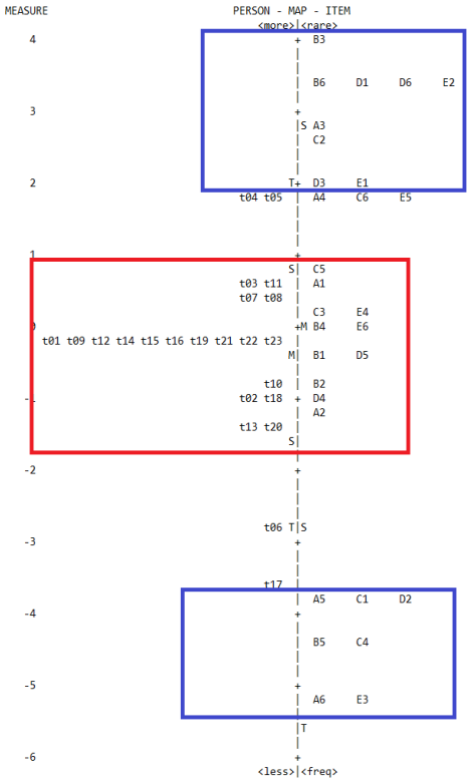
In general, for the England sample, the summary statistics of persons show that the reliability of persons' estimates is not very high but acceptable. It might be due to some of the items being either too easy or too difficult for the sample. Therefore, the identification of different levels of persons' abilities is limited.

Wright Map.

Figure 29 shows that 30 items spread from bottom to top of the scale and covered all the sample abilities. Most of the persons' estimates are located between two 'S' in the red square on the left side of the scale. However, there are 9 items in the top blue square, which are too difficult as nobody gets a score, and 7 items in the bottom blue square, which are too easy as everyone gets a score. It might be because these items are not well designed to distinguish different levels

of persons' abilities. Considering the small English sample size (23 persons), recruiting more people to fill out the survey would potentially increase the accuracy of item estimates.

Figure 29 Wright map of the English sample



Summary of Items and Persons Misfit Statistics of English Sample. For misfit items, as shown in *Table 36*, INFIT MNSQ values range from .32 to 1.70 and OUTFIT MNSQ values range from .10 to 2.69. INFIT ZSTD values range from -2.37 to 1.86, and OUTFIT ZSTD values range from -1.48 to 1.33. Five items (C4, A4, D2, D3, C5) in *Table 37* with either INFIT MNSQ or OUTFIT MNSQ out of the acceptable range. However, none of the INFIT and OUTFIT ZTSD is greater than 2. Moreover, the PTMEASURE-AL CORR values range from .18 to .74. There is no negative correlation which indicates that all the items align with the persons' abilities.

Table 36 Item misfit statistics

Item	INFIT	OUTFIT	PTMEASURE-

MNSQ	ZSTD	MNSQ	ZSTD	AL CORR
.32 - 1.70	-2.37 - 1.86	.10 - 2.69	-1.48 - 1.33	.18 - .74

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

Table 37 The misfit items statistics

Items	INFIT		OUTFIT		PTMEASURE -AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
C4	.79	-.10	2.69	1.33	.29
A4	1.70	1.74	1.85	1.26	.18
D2	.91	.03	1.81	.98	.36
D3	1.57	1.40	1.33	.66	.34
C5	1.56	1.86	1.35	.93	.44

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

For persons, as shown in *Table 38*, INFIT MNSQ values range from .52 to 2.52 and OUTFIT MNSQ values range from .23 to 4.55. INFIT ZSTD values range from -1.28 to 3.65, and OUTFIT ZSTD values range from -.13 to 2.73. There is 1 person with INFIT MNSQ out of acceptable range, and ZSTD value is greater than 2; another person with OUTFIT MNSQ out of acceptable range, and ZSTD value is greater than 2. Therefore, in total, there are only 2 people who are outliers.

Table 38 Person misfit statistics

Person	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
	.52 - 2.52	-1.28 - 3.65	.23 - 4.55	-.13 - 2.73	.57 - .89

Note. MNSQ=Mean Square Statistic, ZSTD=Standard as a Z-score, PTMEASURE-AL CORR= Point Measure Correlation

Dimensionality and Local Dependency. The study investigated the Dimensionality Map in Winsteps to look at the statistical characteristics to see whether there is a sub-dimension of the scale for the English sample. In *Table 39*, the Rasch dimension explains 72.2% of the variance in data which is good. However, the unexplained variance in the first contrast is 5.20 units of Eigenvalue. The first contrast with an eigenvalue bigger than 2 should be concerned (Linacre, 2017). One Eigenvalue unit equals 1 item, which means that in our scale, there are almost 6 items the residuals show a pattern of potential sub-dimension.

Figure 30 shows the loading of the first contrast of items. The X-axis represents the difficulty measures of items, and Y-axis represents the loadings on the conceptual component in the unexplained variance. The positive loading items show a cluster of the largest correlation with the latent sub-dimension (Linacre, 2017). In *Figure 30* the highlighted items A, B, C, D, E, and F have high loadings of over .4. *Table 40* shows the details of items with the largest loading in the first contrast. An examination of the content of these six items shows that these behaviours are from different scenarios, and they are different behaviours. From the content investigation, they do not seem to share characteristics to form a different dimension of the latent the instrument tries to measure.

Table 41 shows the items' largest standardized residual correlations to identify any dependent items. The paired items with high positive correlations indicate local dependency. The paired items D2 and C4 share 86% of the variance in their residuals, which is high. Linacre (2017) recommended that correlations be around 0.7 before we are concerned about dependency. We have five pairs of items whose correlation is higher than .7. In *Table 41,* item E3 and A6 have similar behaviour. They both present concentrated behaviour in different scenarios. E3 presents concentrated behaviour in an inattentive scenario, and A6 presents concentrated behaviour in

a disruptive scenario. Therefore, the correlation between the two similar behaviour is 1. Although item A2 and B2 are from the same scenario, 'Fidgety', the two behaviours are entirely different. The other three pairs of items are from different scenarios and completely different behaviours. There is no correlation between the behaviours.

Furthermore, the English sample size (size=23) is relatively small, and it cannot rule out the possibility that high correlation is caused by noise and chance. Other statistics show that the scale has relatively high reliability for both items and persons. The rating categories work well for the English teacher. Item and person fit for Rasch model. In conclusion, the scale works for English teachers.

Table 39 Standardized Residual Variance

	Eigenvalue	Observed	Expected
Total raw variance in observations	107.94	100%	100%
Raw variance explained by measures	77.94	72.2%	72.4%
Raw variance explained by persons	18.33	17.0%	17.0%
Raw variance explained by items	59.61	55.2%	55.4%
Raw variance unexplained variance (total)	30.00	27.8%	27.6%
Unexplained variance in 1st contrast	5.20	4.8%	17.3%

Figure 30 Standardized residual first contrast plot

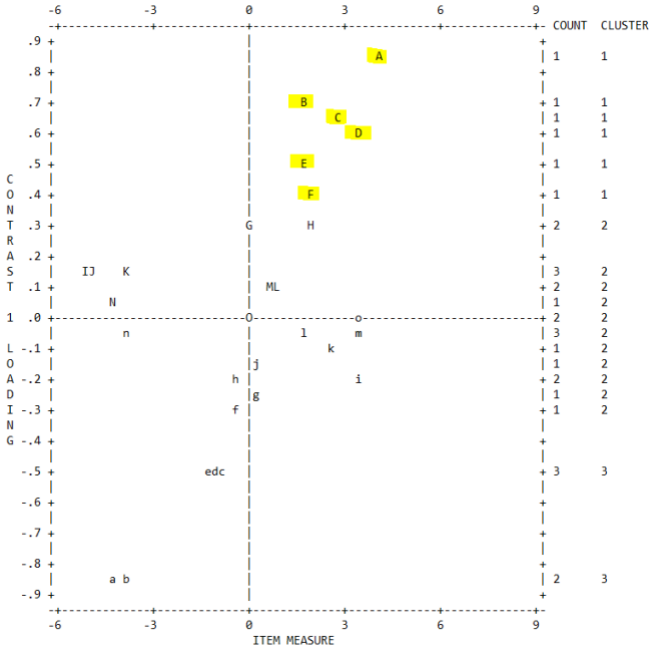


Table 40 Content of 1st contrast items with largest loadings

Label in plot	Label in scale	Loading	Content
A	B3	.87	Tearing paper
B	E5	.68	Leaving seat to chase butterfly
C	A3	.67	Doodling
D	D1	.61	Leaving seat and eating sweets
E	C6	.50	Tickling another child
F	D3	.40	Eating snacks

Table 41 Largest standardized residual correlations used to identify dependent item

Correlation	Item	Item
1.00	E3 Concentrating	A6 Concentrating on writing
.93	D2 Sitting still	C4 Tidy table
.77	D3 Eating snack	C6 Tickling another child
.77	A2 Moving body on seat	B2 Fiddling with a pencil and tapping feet
.70	A4 Late for school	E6 Disturbing the teacher

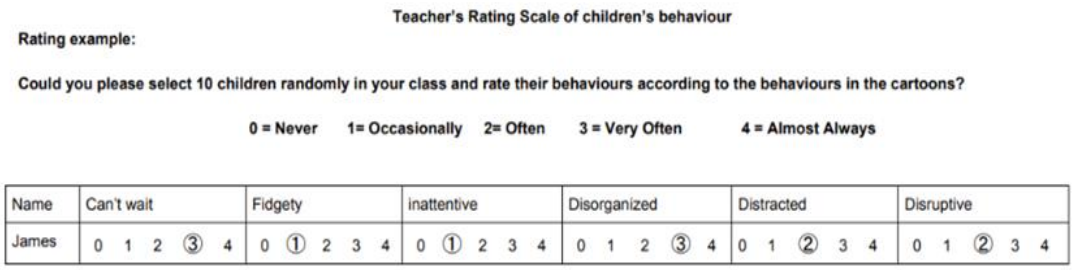
In conclusion, the Rasch Model analysis of the rating scale for the English sample provides valuable insights into its reliability and functionality. The category function analysis, based on INFIT and OUTFIT MNSQ values, suggests that the scale is productive for measuring the sample, falling within the recommended range for rating scale performance. The Andrich Threshold also meets the necessary criteria, indicating a satisfactory advance in step difficulties between categories. Probability curves further support the scale's effective functioning. Summary statistics for persons and items indicate good overall reliability and separation, although some items might need refinement for better discrimination. The Wright Map illustrates a balanced distribution of items and persons along the logit scale, with minor issues related to item difficulty. Item and person misfit analysis identifies a limited number of outliers. Dimensionality analysis raises a potential sub-dimension related to focusing behaviour but not representing a distinct latent trait. Importantly, there is no evidence of dependent items. Despite the relatively small sample size, the scale shows promising reliability and functionality for assessing English teachers.

8.3.3 The Teacher's Rating Scale of Children's Behaviours (TRSCB)

Description

Following the completion of watching the six cartoons and evaluating the animated children's behaviours, teachers proceeded to rate the randomly chosen 10 children from their class. The rating was conducted using a Likert scale ranging from 0 to 4, as shown in *Figure 31*. The scale encompassed six symptoms: "Can't wait," "Fidgety," "Inattentive," "Disorganized," "Distracted," and "Disruptive." Each symptom was assigned a rating from 0 to 4, with 0 representing "Never," 1 indicating "Occasionally," 2 denoting "Often," 3 reflecting "Very Often," and 4 signifying "Almost."

Figure 31 Teacher's Rating Scale of Children's Behaviour



Teacher's Rating Scale of children's behaviour

Rating example:

Could you please select 10 children randomly in your class and rate their behaviours according to the behaviours in the cartoons?

0 = Never 1= Occasionally 2= Often 3 = Very Often 4 = Almost Always

Name	Can't wait	Fidgety	inattentive	Disorganized	Distracted	Disruptive
James	0 1 2 ③ 4	0 ① 2 3 4	0 ① 2 3 4	0 1 2 ③ 4	0 1 ② 3 4	0 1 ② 3 4

Purpose

The rating scale was utilised to evaluate and assess the behaviours of real children in the teachers' classrooms across two countries to conduct comparison study. By randomly selecting 10 children from each teacher's class and having the teachers rate their behaviours against the ADHD symptoms, the study aimed to gather data on the presence and severity of ADHD-related behaviours in an actual classroom setting across two countries. Then the study are able to compare the ratings of children's behaviours between the countries.

Scoring and Interpretation

The 0 to 4 Likert scale used in the Teacher's Rating Scale of Children's Behaviours serves as a quantitative measure to assess the severity or frequency of ADHD-related behaviours exhibited by children in the classroom. Here is a general guide for scoring and interpreting the ratings: 0: Never - The behaviour is never observed or occurs extremely rarely. It indicates that the particular ADHD symptom is not present or has minimal impact on the child's behaviour. 1: Occasionally - The behaviour is observed sporadically or infrequently. It suggests that the ADHD symptom is present, but it occurs only on rare occasions or has a minimal impact. 2: Often - The behaviour occurs frequently or regularly. It indicates that the ADHD symptom is noticeable and manifests relatively frequently in the child's behaviour. 3: Very Often - The behaviour occurs consistently and significantly. It suggests that the ADHD symptom is highly prevalent and consistently affects the child's behaviour. 4: Almost Always - The behaviour is almost constantly present or has a significant impact. It indicates that the ADHD symptom is pervasive and greatly influences the child's behaviour, occurring nearly all the time.

8.3.4 The Reliability and Function of the Teacher's Rating Scale of Children's Behaviours (TRSCB)

To explore the reliability of TRSCB for the Chinese and England sample, the study investigated category function, a summary of statistics, misfit persons and items, unidimensionality and local dependence for both samples separately.

The Reliability of the TRSCB for Chinese Sample

For the Chinese sample, there are 550 children rated by 55 teachers.

Category Function. In order to know whether the categorical scale (0, 1, 2, 3, 4) is productive for measuring the Chinese sample, the study investigated the category function. INFIT and OUTFIT MNSQ from .6 to 1.4 is considered productive for rating scale measurement (Linacre & Wright, 1994). In *Table 42*, the INFIT MNSQ values range from .84 to 1.43 and OUTFIT MNSQ values range from .82 to 1.42. It indicates that the scale is productive for measuring the

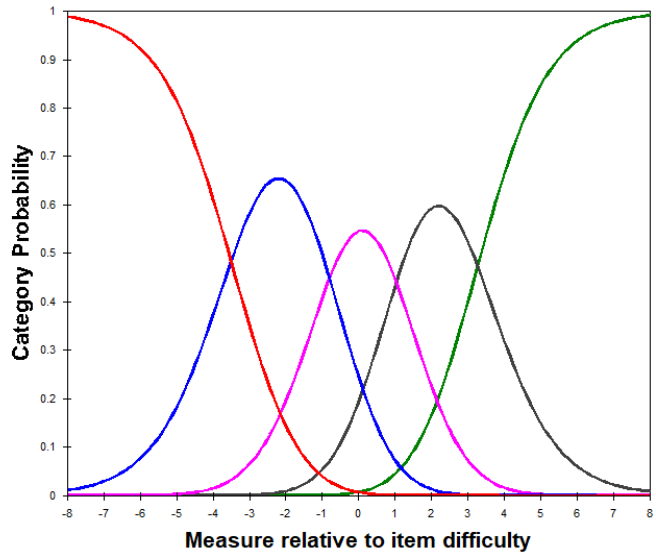
sample. Linacre (1999) recommended a minimum of 1.4 logit for the Andrich threshold. In *Table 42*, the ANDRICH THRESHOLD between categories is about 2 logit. It means that the necessary minimum of advance in step difficulties between these categories is satisfied.

Figure 32 shows that the categorical scale functions reasonably well for measuring the sample. The curves summarise how the sample uses the response (0,1, 2, 3, 4) in the measurement. The horizontal x-axis represents the item's difficulties. The vertical y-axis represents the expected possibility of endorsement of any given category when a teacher gives a rating to an item. The minimum threshold value for the highest probability peak for any category is expected to be.5 in order for the category to be interpreted as functioning reasonably well. In Figure, five peaks represent the probabilities of endorsement of five categories. In general, the category is functioning well in the Chinese sample.

Table 42 Summary of category function of Chinese children sample

Category Label	INFIT MNSQ	OUTFIT MNSQ	Andrich Threshold	Category measures
0	1.23	1.15	None	-4.65
1	.85	.85	-3.50	-2.19
2	.84	.82	-.80	.10
3	.88	.90	1.04	2.20
4	1.43	1.42	3.27	4.44

Figure 32 Category probabilities



Summary Statistics for Chinese Sample. *Table 43* shows the summary statistics of persons' measures from this scale. The average person's measure (Mean) is -.77 logit. The person's measure is in a wide range of -5.28 logit to 5.06 logit with a standard deviation (P.SD) of 2.33 logit for person's estimates, indicating that most of the population spread in a relatively good range of 4.66 logit on the scale. The reliability of the children's ADHD behaviours estimates is .89, and separation is 2.84, which is good. It indicates that the items are reliable and sensitive for distinguishing persons with different levels of abilities.

Table 43 shows the average (Mean) measure of items located at 0 logit as default. The difficulty of items ranges from -.91 logit to .83 logit. Most item difficulties are located in a standard deviation of .63 logit. It indicates that items spread in a narrow range of difficulties to measure the trait. Item reliability and separation indicate the ability of the measurement to define a distinct hierarchy of items along the measured trait. The higher the number, the more confident the study can apply the measurement across other samples and get the same order of item estimates. The item reliability is .99, and the separation is 8.56. They both are high. It suggests that the study can rely on this order of item estimates for other suitable samples.

Table 43 Summary statistics of persons and items

	Mean	P.SD	Min	Max	Separation (Real RMSE)	Person/Item Reliability
Persons	-.77	2.33	-5.28	5.06	2.84	.89
Items	.00	.63	-.91	.83	8.56	.99

Persons and Items Misfit Statistics. *Table 44* shows that the items' INFIT MNSQ values range from .81 to 1.29, and OUTFIT MNSQ values range from .79 to 1.22. INFIT ZSTD values range from -3.37 to 4.24, and OUTFIT ZSTD values range from -3.56 to 3.35. All the INFIT MNSQ are in the good range (.60 to 1.4). Moreover, the PTMEASURE-AL CORR values range from .82 to .87. All the items have strong correlations, and there is no negative correlation which indicates that all the items align with the persons' abilities.

Table 44 Item misfit statistics

Item	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
Can’t wait	1.29	4.24	1.22	3.35	.82
Disorganised	1.16	2.45	1.1.	1.57	.82
Disruptive	1.09	1.33	1.10	1.41	.82
Fidgety	.85	-2.50	.83	-2.96	.87
Distracted	.81	-3.38	.82	-3.04	.87
Inattentive	.81	-3.37	.79	-3.56	.87

Table 45 shows that persons' INFIT MNSQ values range from .09 to 6.29, and OUTFIT MNSQ values range from .09 to 6.40. INFIT ZSTD values range from -2.81 to 4.84, and OUTFIT ZSTD values range from -2.80 to 4.85.

Table 45 Person misfit statistics

Person	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
	.09 - 6.29	-2.81 - 4.84	.09 - 6.40	-2.80 - 4.85	-.83 - 0

Dimensionality and Local Dependency. To investigate whether the Teacher's Rating Scale of Children's Behaviour is unidimensional, Rasch PCA of residuals is looking for any meaningful sub-dimension in the residuals. *Table 46* Standardized Residual Variance shows that the measure explains 71.1% of the variance in data which is good. Meanwhile, the unexplained variance in the first contrast is 1.90 units of Eigenvalue. The first contrast with an eigenvalue bigger than 2 should be concerned (Linacre, 2017). Therefore, there is no concern.

Figure 33 shows the loading of the first contrast of items. The X-axis represents the difficulty measures of items, and Y-axis represents the loadings on the conceptual component in the unexplained variance. The positive loading items show a cluster of the largest correlation with the latent sub-dimension (Linacre, 2017). In *Figure 33***,** the highlighted items 3 and 5 have high loadings of over .5. *Table 47* shows the details of the first contrast of items with the largest loading. An examination of the content of the two items shows that item 3, 'Inattentive' and item 5, 'Distracted', are two different symptoms of ADHD. Given that the eigenvalue of the first contrast is rather small, the two items do not form a sub-dimension of the latent.

Table 48 shows the items' largest standardized residual correlations to identify any dependent items. The paired items with high positive correlations indicate local dependency. Linacre (2017) recommended that correlations be around 0.7 before we are concerned about dependency. The highest positive correlation in *Table 48* is .23, which means that item

Inattentive and item Distracted only share about 5% of the variance in their residuals in common, which is very small. Therefore, there is no local dependent item in the scale.

Table 46 Standardized residual variance

	Eigenvalue	Observed	Expected
Total raw variance in observations	20.79	100%	100%
Raw variance explained by measures	14.79	71.1%	71.1%
Raw variance explained by persons	11.62	55.9%	55.8%
Raw variance explained by items	3.17	15.2%	15.2%
Raw variance unexplained variance (total)	6.00	28.9%	28.9%
Unexplained variance in 1st contrast	1.90	9.1%	31.7%

Figure 33 Standardized residual first contrast plot

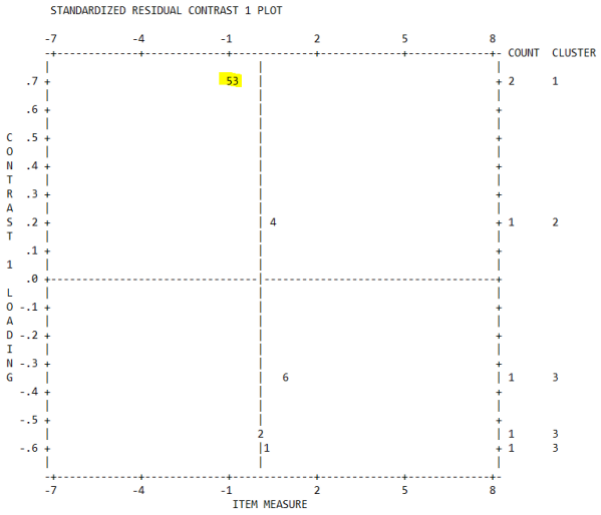


Table 47 Standardized residual first contrast plot

Label in plot	Label in scale	Loading
3	Inattentive	.72
5	Distracted	.72
4	Disorganised	.21
1	Can't wait	-.62
2	Fidgety	-.55
6	Disruptive	-.37

Table 48 Largest standardized residual correlations used to identify dependent item

Correlation	Item	Item
.23	Inattentive	Distracted
.03	Can't wait	Fidgety

In conclusion, the Rasch Model analysis conducted separately for the Chinese samples provides a comprehensive assessment of the reliability and functionality of the Teacher's Rating Scale of Children's Behaviour (TRSCB). For the Chinese sample, consisting of 550 children rated by 55 teachers, the analysis indicates that the categorical scale (0, 1, 2, 3, 4) effectively measures the sample, as evidenced by INFIT and OUTFIT MNSQ values falling within the recommended range for rating scale performance. The Andrich Threshold also meets the necessary criteria, ensuring a suitable advance in step difficulties between categories. Probability curves further affirm the scale's functionality for this sample. Summary statistics for persons and items demonstrate good reliability, with the scale proving sensitive in distinguishing individuals with varying levels of abilities. Misfit statistics for both items and persons are generally within acceptable limits, indicating that the items align well with the abilities of the respondents. Additionally, the scale's unidimensionality is supported by Rasch PCA of residuals, with a high percentage of variance explained by the measure and no concerning eigenvalues in the first contrast. In summary, the TRSCB exhibits strong reliability and functionality for both the Chinese, making it a valuable tool for assessing children's behaviour in educational contexts.

The Reliability and Function of the TRSCB for English Sample

There are 210 children rated by 21 teachers in England.

Category Function. In order to know whether the categorical scale (0, 1, 2, 3, 4) is productive for measuring the English sample, the study investigated the category function. INFIT and OUTFIT MNSQ from .6 to 1.4 is considered to be productive for rating scale measurement (Linacre & Wright, 1994). In *Table 49,* the INFIT MNSQ values range from .84 to 1.34 and OUTFIT MNSQ values range from .79 to 1.28. It indicates that the scale is productive for measuring the sample. Linacre (1999) recommended a minimum of 1.4 logit for the Andrich threshold. In *Table 49,* the Andrich Threshold between categories' 2' (Often) and '3' (Very

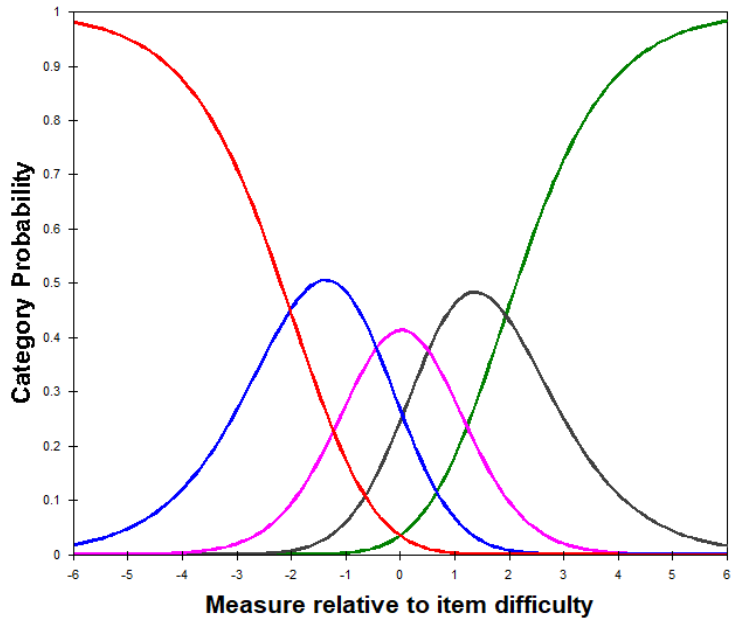
Often) is .95, which is lower than the minimum 1.4 logit. It indicates that English teachers might not differentiate between 'Often' and 'Very Often' very well.

In *Figure 34,* the probability curves show that the categorical scale functions reasonably well for measuring the sample. The curves summarise how the sample uses the response (0,1, 2, 3, 4) in the measurement. The horizontal x-axis represents the item's difficulties. The vertical y-axis represents the expected possibility of endorsement of any given category when a teacher gives a rating to an item. The minimum threshold value for the highest probability peak for any category is expected to be .5 in order for the category to be interpreted as functioning reasonably well. In *Figure 34*, the space between the peaks of '2' and '3' is very narrow, and the expected possibility of endorsement of category '2' is lower than .5. This information indicates that English teachers might not interpret the difference between category 'Often' and 'Very Often' very clearly. However, five peaks represent the probabilities of endorsement of five categories, and other categories work well for the sample. In general, the category is functioning satisfactorily for the English sample.

Table 49 Summary of category function of English children sample

Category Label	INFIT MNSQ	OUTFIT MNSQ	Andrich Threshold	Category measures
0	1.03	1.02	None	-3.25
1	.88	.90	-2.02	-1.37
2	.84	.79	-.43	.03
3	.95	1.04	.52	1.37
4	1.34	1.28	1.94	3.19

Figure 34 Category probabilities



Summary Statistics for English Sample. The summary statistics are in *Table 50*. The average person's measure (Mean) is -.70 logit. The person's measure is in a wide range of -3.75 logit to 3.69 logit with a standard deviation (P.SD) of 1.72 logit, indicating that most of the population spread in a good range of 3.44 logit on the scale. The reliability of the teachers' attitude estimates is .84, and the separation is2.32, which is not high but acceptable. It might be due to the limited number of items to differentiate a relatively large sample.

In *Table 50,* the average (MEAN MEASURE) measure of items is located at 0 logit as default. The difficulty of items is in a wide range from -.59 logit to .59 logit. Most item difficulties are located in a standard deviation of .35 logit. It indicates that items spread in a narrow range of difficulties to measure the trait. Item reliability and separation indicate the ability of the measurement to define a distinct hierarchy of items along the measured trait. The higher the number, the more confident the study can apply the measurement across other samples and get the same order of item estimates. The item reliability is .91, and the separation is3.11. They both are high. It suggests the study can rely on this order of item estimates for other suitable samples.

Table 50 Summary statistics for persons and items

	Mean	P.SD	Min	Max	Separation	Reliability
Persons	-.70	1.72	-3.75	3.69	2.32	.84
Items	.00	.35	-.59	.59	3.11	.91

Persons and Items Misfit Statistics. In *Table 51***,** the item misfit statistics, INFIT MNSQ values range from .64 to 1.69 and OUTFIT MNSQ values range from .67 to 1.77. INFIT ZSTD values range from -3.79 to 5.35, and OUTFIT ZSTD values range from -3.38 to 5.49. There is only one item (Can't wait); the MNSQ values are slightly out of range, and the rest items are in a good range (0.6-1.4). Moreover, the PTMEASURE-AL CORR values range from .80 to .84. All the items have a strong correlation, and there is no negative correlation which indicates that all the items align with the persons' abilities.

Table 51 Item misfit statistics

Item	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
Can’t wait	1.69	5.35	1.77	5.49	.82
Disorganised	1.36	3.06	1.35	2.84	.83
Disruptive	.93	-.64	.81	-1.62	.80
Inattentive	.74	-2.67	.71	-2.86	.82
Distracted	.65	-3.79	.67	-3.38	.84
Fidgety	.64	-3.83	.60	-4.19	.83

In *Table 52,* person misfit statistics, INFIT MNSQ values range from .07 to 4.0 and OUTFIT MNSQ values range from .07 to 3.93. INFIT ZSTD values range from -2.95 to 3.48 and OUTFIT ZSTD values range from -2.92 to 3.36.

Table 52 Person misfit statistics

Person	INFIT		OUTFIT		PTMEASURE-AL CORR
	MNSQ	ZSTD	MNSQ	ZSTD	
	.07 - 4.0	-2.95 - 3.48	.07 - 3.93	-2.92 - 3.36	-.83 - 0

Dimensionality and Local Dependency. To investigate whether the scale of Teacher's Rating Scale of Children's Behaviour is unidimensional, Rasch PCA of residuals is looking for any meaningful sub-dimension in the residuals. In *Table 53,* Standardized Residual Variance shows that the measure explains 64.9% of the variance in data which is good. Meanwhile the unexplained variance in 1st contrast is 1.86 units of Eigenvalue. The fist contrast with an eigenvalue bigger than 2 should be concern (Linacre, 2017).

Figure 35 Standardized residual first contrast plot shows the loading of the first contrast of items. The X-axis represents the difficulty measures of items, and Y-axis represents the loadings on the conceptual component in the unexplained variance. The positive loading items show a cluster of the largest correlation with the latent sub-dimension (Linacre, 2017). In *Figure 35,* the highlighted items 1 and 6 have high loadings of over .5. *Table 54* shows the details of items with the largest loading. An examination of the content of the two items shows that item 1, 'Can't wait', and item 6, 'Disruptive', are two different symptoms of ADHD. Given that the eigenvalue of the first contrast is relatively small, the two items do not form a sub-dimension of the latent.

Table 55 shows the items' largest standardized residual correlations to identify any dependent items. The paired items with high positive correlations indicate local dependency. Linacre (2017) recommended that correlations be around 0.7 before we are concerned about dependency. The highest positive correlation in *Table 55* is .09, which means that item Inattentive and item Distracted only share about 1% of the variance in their residuals in common, which is very small. Therefore, there is no local dependent item in the scale.

Table 53 Standardized Residual Variance

	Eigenvalue	Observed	Expected
Total raw variance in observations	17.07	100%	100%
Raw variance explained by measures	11.07	64.9%	64.8%
Raw variance explained by persons	7.54	44.2%	44.1%
Raw variance explained by items	3.53	20.7%	20.7%
Raw variance unexplained variance (total)	6.00	35.1%	35.2%
Unexplained variance in 1st contrast	1.86	10.9%	31.0%

Figure 35 Standardized residual first contrast plot

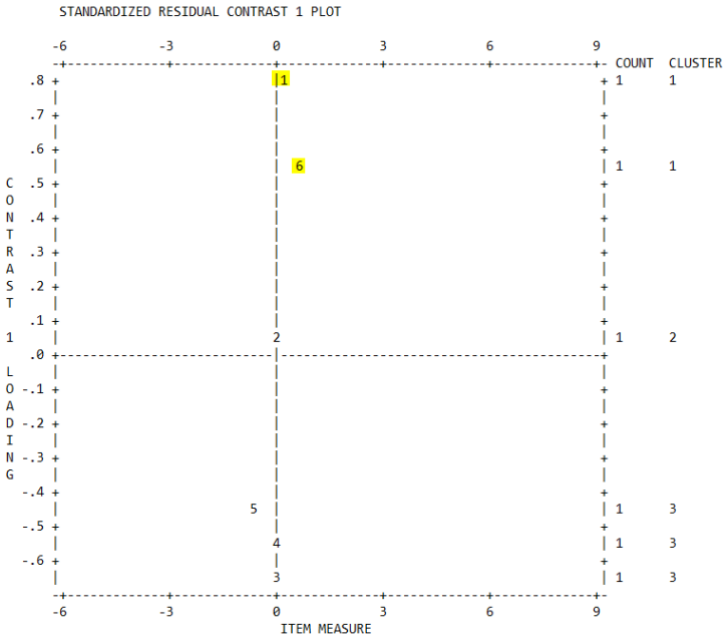


Table 54 Content of 1st contrast items with largest loadings

Label in plot	Label in scale	Loading
1	Can't wait	.78
6	Disruptive	.57
2	Fidgety	.05
3	Inattentive	-.65

4	Disorganised	-.56
5	Distracted	-.43

Table 55 Largest standardized residual correlations used to identify dependent item

Correlation	Item	Item
.09	Inattentive	Distracted
.01	Can’t wait	Disruptive
.01	Inattentive	Disorganised

In conclusion, the Rasch Model analysis conducted for the English sample provides valuable insights into the reliability and functionality of the Teacher's Rating Scale of Children's Behaviour (TRSCB). For the English sample, which comprises 210 children rated by 21 teachers, the analysis reveals several important findings. First, the categorical scale (0, 1, 2, 3, 4) effectively measures the sample, as indicated by INFIT and OUTFIT MNSQ values falling within the recommended range for rating scale performance. However, there is a minor issue with the differentiation between the categories 'Often' and 'Very Often,' as suggested by the lower Andrich Threshold between these categories. Probability curves confirm the scale's overall functionality, with most categories performing well. Nevertheless, there is a slight ambiguity in distinguishing between 'Often' and 'Very Often,' as evidenced by a narrow gap between the corresponding probability peaks.

Summary statistics for the English sample show that the scale's reliability for teacher ratings is acceptable, with a reliability coefficient of .84. However, the separation value of 2.32 suggests some limitations in differentiating among a relatively large sample, possibly due to the limited number of items. Item analysis reveals that item difficulties are concentrated within a narrow range, but the scale still demonstrates high item reliability and separation, indicating consistent item hierarchy across different samples. There are no negative correlations between items and persons, indicating good alignment with respondents' abilities. The assessment of

dimensionality through Rasch PCA of residuals demonstrates that the scale is predominantly unidimensional, with a high percentage of variance explained by the measure and no concerning eigenvalues in the first contrast. Moreover, there is no evidence of local item dependency, further supporting the unidimensional nature of the scale.

In summary, the Teacher's Rating Scale of Children's Behaviour (TRSCB) exhibits strong reliability and functionality for the English sample, with minor issues related to the differentiation of certain categories. Despite these minor concerns, the scale proves to be a valuable tool for assessing children's behaviour in an English educational context, particularly for identifying symptoms of ADHD.

8.4 Sampling method

8.4.1 Schools

In China, children aged 6 to 15 years are entitled to a free place at a public school. Primary education begins at age 6 and continues till age 12 (Compulsory Education Law of the People's Republic of China, 2006). In England, children aged 5 to 16 years are entitled to a free place at a state school where primary education begins at age 5 and continues till age 11 ("Types of school", 2022). Almost all children start reception class at school when they are 4 on 1st September. In China, 11 schools participated in the research; 109 teachers and 600 children participated. In England, 15 schools participated in the study; 24 teachers and 240 children participated. In both countries, the research focussed on Year 2 (Y2).

In China, there are only two types of schools, public schools and private schools (Ministry of Education of the People's Republic of China Government, 2020). The public schools are entirely funded by the government and follow the national curriculum Ministry of Education of the People's Republic of China Government, 1996). However, private schools are self-

funded and do not have to follow the national curriculum. The children who attend private schools must pay a tuition fee. There are broadly two types of schools in England: state and private schools. Children attending private schools must pay a fee. There are many types of state schools, such as community, foundation, voluntary, academies, and grammar schools ("Types of school", 2022). The main differences among these schools are runners, freedom of choosing what they teach and student recruitment. *Figure 36* and *Figure 37* shows the percentages of each type of school in our sample.

Figure 36 School types in the Chinese sample

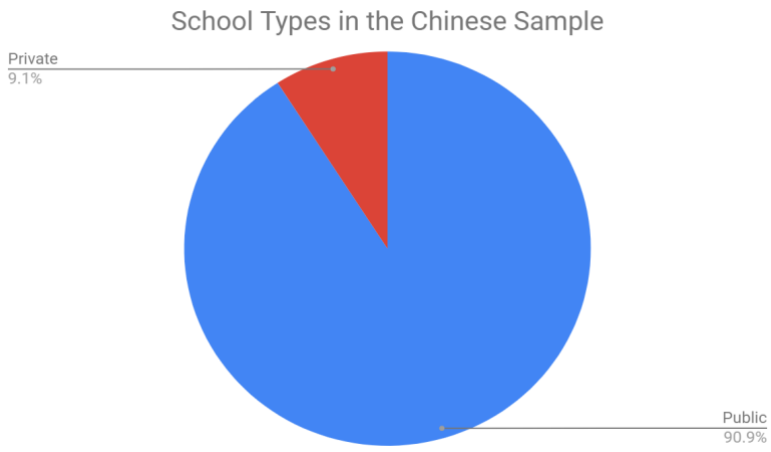
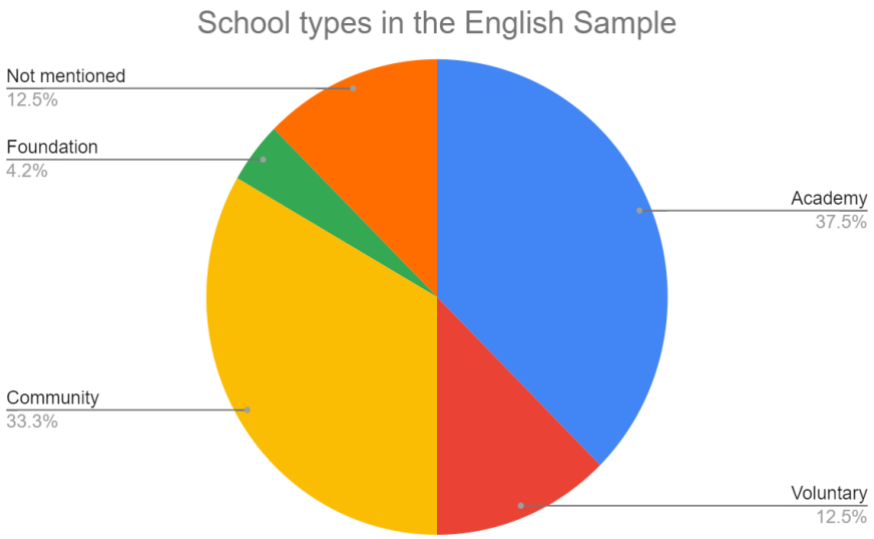


Figure 37 School types in the English sample



In the Chinese sample, only one private school participated in the research. The participating schools have at least two Year 2 (Y2) (aged 7 to 8 years) classes; some even have 11 Y2 classes. There is no class size legislation in China, and the class size varies from 32 to 60 pupils in one class. The average class size is 47. In England, the legislation class size is 30 pupils in one class (Department of Education, 2011). In this study, the smallest class size is 10, and the largest class size is 30 in the English sample. The average class size is 23.

8.4.2 Teachers

In China, the national curriculum of Y2 includes six subjects (Chinese, English, Math, Music, PE, Art/Craft) (Ministry of Education of the People's Republic of China Government, 2001). A subject teacher teaches each subject. Therefore, each class is taught by six or five teachers, and one of them is the class teacher who is in charge of teaching not only a subject but also parent meetings, registration and school report. One subject teacher teaches more than one Y2 class. In the Chinese sample, 109 teachers participated in the study. The characteristics of gender, teaching years (TY), class teacher (CT), and subject specialist (SS) are shown in *Table 56* and *Table 57* below.

Table 56 The characteristics of the Chinese teacher sample

Participants	Female	Male	Max TY	Min TY	Average TY	CT	SS
109	97	12	36	1	16.09	61	48

Note. Max TY=Maximum Teaching Years; Min TY= Minimum Teaching Years; Average TY=Average Teaching Years; CT=Class Teacher; SS= Subject Specialist

Table 57 The number of subject specialists in each subject

Subjects	Chinese	Math	Art/Craft	Music	PE	English
Number	54	32	4	3	2	14

In England, the national curriculum includes 12 subjects (English, Mathematics, Science, Art & Design, Computing, Design & Tech, Geography, History, Languages, Music, Physical Ed) ("National curriculum in England: a framework for key stages 1 to 4", 2014). A Y2 teacher teaches all the subjects. Some schools have a teaching assistant in the Y2 class to help with the teaching activities, while others only have a class teacher. In the English sample, 24 teachers participated in the study. The characteristics of gender, teaching years (TY), class teacher (CT), and teaching assistant (TA) are shown in *Table 58.*

Table 58 The characteristics of the English teacher sample

Participants	Female	Male	Max TY	Min TY	Average TY	CT	TA
24	21	3	20	1	8.63	15	8

Note. Max TY=Maximum Teaching Years; Min TY=Minimum Teaching Years; Average TY=Average Teaching Years; CT=Class Teacher; TA=Teaching Assistant.

8.4.3 Children

In total, 840 children were selected randomly and rated by their teachers in the study. In the Chinese sample, there were 600 children: 282 girls and 318 boys. In the English sample, there

were 240 children: 114 girls and 126 boys. The Chinese children were selected randomly by the researcher. The English children were selected randomly by their teachers.

8.5 Data Collection

8.5.1 China

In China, before carrying out data collection, I contacted the Education Department of the local city to have their permission to conduct the study at schools and their support to recruit schools. All 11 schools are the schools which have agreed to participate. The data collection was carried out at the beginning of Year 2 new semester in September and finished at the end of September. Although it is a new semester, all the teachers who participated in my study had taught the pupils since Year 1 when the children had just started their compulsory education at school. Therefore, they knew the children very well. I have included almost all the teachers from the participating schools except a few teachers who were unavailable.

The 11 schools have 60 Year 2 classes in total. I randomly selected 10 children from each class and asked 2 to 5 teachers who taught this class to rate the children's ADHD behaviours and the cartoon characters. In order to generate a random sample from the population of participating schools, I used Microsoft Excel to generate a random sample from each class. First, generate a random value for each child in the class with Excel Function RAND (). Second, Sort the order of all the children with the random value. Third, choose the top 10 children from the order.

The questionnaire packages include the consent form, the questionnaires, a thank card and a small English present to thank the teachers for participating in the study. Since I had permission from the Education Department and schools, I was able to go to the participating schools and stay at the schools to hand out the questionnaires to teachers and collect them after they finished. In participating schools, not every teacher has a computer at work. In most schools, 6 or 8

teachers share one computer, and some computers have no internet connection. Therefore, I brought my laptop to the school and played the cartoon animations for teachers to rate. I was introduced to the Year 2 teachers by the head teachers. I stayed at the teacher's office. When they had time, they came to me to collect the questionnaires and returned them after they finished. So, the questionnaire return rate is 100% in China. In total, 11 schools, 60 Year 2 classes, 109 Year 2 teachers and 600 children participated in the study in China.

8.5.2 England

In England, the participating teachers were recruited in two ways. One is with the help of the School of Education at Durham University. With my supervisor's introduction, I got in touch with the organiser of the schoolteacher training programme. The school has meetings with the Headteachers, which have teachers participating in the teachers' training program. I was given chances to present my study to the headteachers and ask them to pass the questionnaires to their Year 2 teachers or teaching assistants to fill out. Another way is asking my friends and colleagues who either themselves are Year 2 teachers or have family members or friends who are Year 2 teachers to take part in the study. The questionnaire packages were handed to my friends or colleagues to fill out or pass to their family members or friends to fill out. Some were handed out by post.

The questionnaire packages include the consent form, the questionnaires, the link to the six cartoons, an envelope with a return address and stamp, a thank card and a small Chinese present to thank the teachers for participating in the study. All six cartoons were uploaded on my YouTube channel. However, the return rate is relatively low in England. I have sent out 75 questionnaires. I have 24 questionnaires returned to me. It also took me longer (8 months) to collect data in England. I think the main reason that the return rate is very low (32%) is that I did not have direct contact with the teachers who have agreed to take part. Therefore, it is hard

to follow up after they have the questionnaires. I have tried to recruit more teachers by emailing schools to ask them to participate in the study. However, I did not get any replies from the schools. I have also tried to knock on the doors of schools. However, due to the safety and security reasons of the schools in England, I could not enter the schools. Although I talked to the receptions, I was told that the Year 2 teachers and teaching assistants were very busy in the class. In total, 15 schools, 24 teachers and 240 children participated in the study in England.

8.6 Analysis Strategy

Prior to data analysis, the collected ratings were subjected to data cleaning procedures. This involved checking for any missing or incomplete data and identifying potential outliers. This analysis was conducted using software SPSS.

Research Question 1: How to measure the leniency of Chinese and English teachers in rating children's ADHD behaviours?

Data collected from the Teacher's Rating Scale of Cartoons (TRSC) was subjected to a descriptive analysis to explore the characteristics of teachers' ratings from both China and England. Additionally, a comparative analysis was conducted to identify significant differences in teachers' ratings between the two countries for each ADHD behaviour. The potential influence of factors such as gender, teaching experience, teaching position, and class size on teachers' ratings was also investigated. SPSS was used for these analyses, employing Descriptive Analysis, Independent Sample T-Test, Pearson correlations, and One-Way ANOVA ($P < 0.05$, two-tailed).

Subsequently, the data from the Teacher's Rating Scale of Cartoon (TRSC) was transformed into ordinal data (0, 1, 2, 3) for Rasch Model analysis. This analysis has three steps. The first

step was to assess the reliability and functioning of the TRSC instrument in measuring the leniency of Chinese and English teachers separately. Various statistical indexes, such as Categorical Function, Wright Map, Summary Statistics, Items and Persons Misfit, Dimensionality, and Local Dependency, were utilised to evaluate the instrument's reliability and functionality. The second step was to calibrate the instrument for measuring teacher's leniency equally across both samples (Chinese and English teachers). It was performed through a Differential Item Functioning (DIF) analysis to exclude behaviours that did not measure teachers' leniency equally across the two countries. Misfit items and persons were also excluded from the analysis to generate accurate measures. Finally, the Rasch Model analysis conducted with Winsteps software yielded Chinese and English teachers' leniency measures on an equal-interval scale.

Research Question 2: How to make appropriate adjustments to ratings of children's ADHD behaviours to improve fairness in cross-cultural comparison?

The data from the Teacher's Rating Scale of Children's Behaviours (TRSCB) underwent analysis using the Rasch Model and Many-facet Rasch Model. This analysis involved three key steps. Firstly, the reliability and functioning of the TRSCB instrument in measuring Chinese and English children's ADHD behaviours were examined separately for each sample. Secondly, the instrument was calibrated to measure children from both countries equally on one interval scale, taking into account the variables of country and gender through a Differential Item Functioning (DIF) analysis. Both of these steps were carried out using Winsteps software. Lastly, teachers' leniency measures from the TRSC instrument were incorporated as anchored values in the Many-facet Rasch Model analysis of the TRSCB instrument. This step allowed for the adjustment of children's ADHD behaviour measures,

resulting in fair scores for Chinese and English children on an equal-interval scale for further analysis. The Many-facet Rasch Model analysis was conducted using Facets software.

Research Question 3: To what extent do the cultural differences affect the ratings of children's ADHD behaviours between England and China?

To address this research question, changes in children's measures after adjusting for teachers' leniency were investigated. This analysis, conducted with SPSS software, involved Paired Sample T-Tests to compare children's measures before and after adjusting for teachers' leniency, determining whether significant changes occurred. Additionally, a Two-Way ANOVA was performed to examine the interaction effects of country and gender on children's ratings after adjusting for teachers' leniency.

Research Question 4: Comparability of ADHD Behaviour Ratings Among Chinese Teachers

400 Chinese children were cross rated by their respective teachers using the TRSCB instrument. The data were subjected to Intraclass Correlation Coefficient (ICC) analysis with the two-way-mixed model using SPSS software. This analysis aimed to determine the inter-rater agreement among Chinese teachers. The average ICC estimates were calculated for each group of cross-rating teachers and each ADHD symptom to assess the level of agreement among teachers in their ratings.

8.7 Conclusion

In this chapter, the researcher has presented a comprehensive overview of the methods and methodologies employed in this study, aiming to compare ADHD behaviours as rated by teachers in China and England. The research was thoughtfully designed to address research

gaps regarding measuring teachers' leniency in a cross-cultural context and adjusting children's ADHD behavioural ratings to ensure a fair comparison between the two countries.

Firstly, this chapter thoroughly explained the sampling method and data collection employed in the study, which was convenience sampling. This approach was chosen for practical reasons, as it allowed access to a suitable number of participants from both China and England. Furthermore, this chapter provided detailed characteristics of the samples, including information about the teachers and children involved in the study. Understanding the demographics of the participants was crucial for interpreting the results accurately and understanding potential biases.

This chapter also described the measurements used in the data collection process, specifically the Teacher's Rating Scale of Cartoons (TRSC) and the Teacher's Rating Scale of Children's Behaviours (TRSCB). Each instrument was clearly described, outlining its purpose, explaining the scoring system, and clarifying how to interpret the results. These measurements served as the foundation for the study's analysis. Additionally, the chapter detailed the evaluation of the reliability of the two instruments through a series of statistical indices from the Rasch Model analysis, ensuring that the instruments provided accurate and consistent measurements for this study. The data analysis strategy was then explained, encompassing various techniques and procedures used to examine the collected data from both instruments. Descriptive analysis, Independent Sample-T Test, Pearson correlations, One-Way ANOVA, and Many-facet Rasch Model were among the methods used to analyse the data and draw meaningful insights.

Overall, the combination of quantitative analyses and systematic data interpretations strengthened the robustness and relevance of the chosen methods and methodologies. These

analyses paved the way for the subsequent chapter, where further analysis was conducted, and results were generated to answer the research questions. By employing these rigorous methods and methodologies, the study aimed to ensure the validity and accuracy of its findings.

Chapter 9 Results

9.1 Introduction

In this chapter, I present the findings of this comprehensive research study, which focuses on the development of measurement instruments and the exploration of rater bias in assessing Attention-Deficit/Hyperactivity Disorder (ADHD) behaviours among teachers from different cultural backgrounds in China and England. The primary objective was to address the challenges arising from subjective judgments of children's ADHD behaviours, influenced by diverse cultural and social contexts. To achieve this, I employed a multi-method approach, involving the creation of standardised cartoon animations as vignettes, the development of the Teacher's Rating Scale of Cartoons (TRSC), and the Teacher's Rating Scale of Children's Behaviours (TRSCB), to create culturally unbiased measurement instruments for evaluating teachers' leniency towards children's ADHD behaviours.

To ensure consistency and fairness in the measurements and to account for rater bias, the study employed the Rasch Model analysis and Many-facet Rasch Model. These analytical tools allowed us to establish a robust framework for making meaningful cross-cultural comparisons in ADHD assessments. Additionally, I utilised software such as Winsteps, Facets, and SPSS for data analysis. The data analysis process involved a systematic approach, beginning with the identification of missing and noisy data, followed by a descriptive analysis of the raw data, calibration of instruments for cross-cultural measurements, and the adjustment of children's measures based on teachers' leniency. This rigorous process culminated in comprehensive comparison results that shed light on the complexities of ADHD assessments across different cultural settings.

9.2 Identifying Missing Data and Noisy Data

After collecting the data and extracting them from the questionnaires, data cleaning is the next step before commencing the analysis. Research data undergo repeated stages of entry into information carriers, extraction, transfer to other carriers, editing, selection, transformation, summarisation, and presentation. It is imperative to recognise that errors can arise at any point in the data flow. Data cleaning involves assessing and enhancing data quality, a crucial step for robust analysis.

Data cleaning entails identifying incomplete, incorrect, inaccurate, or irrelevant segments of the data and then making replacements, modifications, or deletions as necessary (Van den Broeck et al., 2005). Missing data and inconsistent data are primary targets for data cleaning. Missing data frequently occur in questionnaire data, stemming from respondents leaving answers blank for various reasons or researchers erroneously transferring data from questionnaires to data forms. Missing data can lead to numerous issues, including diminished statistical power, reduced sample representativeness, biased parameter estimation, and complexity in study analysis. These problems can jeopardise research validity and potentially yield erroneous conclusions (Kang, 2013). Therefore, identifying missing data, understanding their potential causes, and implementing reasonable strategies for handling them are crucial initial steps in data analysis.

Inconsistent or noisy data is characterised by inaccuracies, inconsistencies, and incompleteness due to errors within the dataset (Ridzuan & Wan Zainon, 2019). Such issues often arise when respondents input correct values into incorrect fields, provide illegible answers, make writing errors, or provide responses outside the expected range. Data cleaning is essential to address these noisy data, as they have the potential to mislead data analysis and influence results. The

most effective approach to handling missing and noisy data is prevention, which hinges on careful questionnaire design and meticulous data collection. A well-designed questionnaire that is easy for respondents to comprehend and complete can reduce the likelihood of missing data and response errors, as discussed in the Chapter of Development of Measurement Instruments.

A rigorous data collection process ensures the validity and reliability of research data. However, missing and noisy data are common challenges in the social sciences. Therefore, having appropriate strategies for dealing with them is the first step in data analysis. For handling missing data, the initial step is to determine the extent of missing data, identify variables with missing data, and classify the type of missing data. A common approach for addressing missing data is to exclude cases with missing values from the analysis. If the sample size is sufficiently large and missing cases are random, their impact on analytical power may be insignificant. However, when missing data exhibit non-random patterns that can significantly affect the analysis, imputation methods become necessary.

It is worth noting that the software Winsteps does not mandate complete data for making estimates within the Rasch measurement framework (Linacre, 2021). In Winsteps, missing data primarily matters when it reduces the statistical information available for measurement construction and quality control. Moreover, if missing data distorts measurements (e.g. coding a wrong response as missing), it can indeed impact the analysis. Otherwise, missing data does not affect estimation. Winsteps employs a flexible estimation method to construct measures from non-missing observations while missing data are skipped. When needed, Winsteps can calculate expected values for all observations (present or missing) for which item and person estimates are known (Linacre, 2021). Consequently, if the quantity of missing data has no

substantial impact on statistical power and the missing data occurs randomly, retaining cases with missing values for analysis is a viable option.

In this study, the presence of missing values in the data is minimal. The Chinese sample exhibits 5 missing values in 5 different cases, representing only 0.15% of all values. In the English sample, one case has 3 missing values, constituting just 0.48% of all values. Whether Winsteps skips missing data or estimates expected values for them does not compromise the representativeness of the sample.

To identify noisy data within the questionnaires, the study aimed to pinpoint questionnaires completed by teachers who may not have grasped the rating scale. These teachers, perhaps misunderstanding the scale, provided ratings that contradicted those of their peers. They either consistently rated all behaviours very low or very high, with no differentiation in their ratings. In essence, the study sought to identify teachers who had used the rating scale differently and potentially inappropriately. Cluster analysis, a widely employed multivariate method for examining data behaviour (Aggarwal, 2014), was employed. Cluster analysis categorises objects based on their internal similarity or dissimilarity (Everitt & Hothorn, 2006), grouping similar objects and segregating dissimilar ones. Moreover, cluster analysis is suitable for small datasets and data that do not conform to assumptions such as linearity, normality, and homoscedasticity (Norusis, 2006). Thus, the study utilised cluster analysis to investigate outliers.

9.2.1 Missing Data and Noisy Data in the Chinese Sample for TRSC

In *Table 59,* 4 variables (A4, C6, B2, C1) have 5 missing values in 5 cases in total. *Figure 38* shows that the percentage of missing data (variables 13.33%, cases 4.59%, values 0.15%) is very low. No systematic missing data or patterns have been found in the missing data.

Therefore, cases with missing values do not affect the power analysis with the rest of the data available. The missing data cases are t013, t102, t001, t098, and t030. the missing data cases were kept for further analysis.

Figure 38 Overall summary of missing values

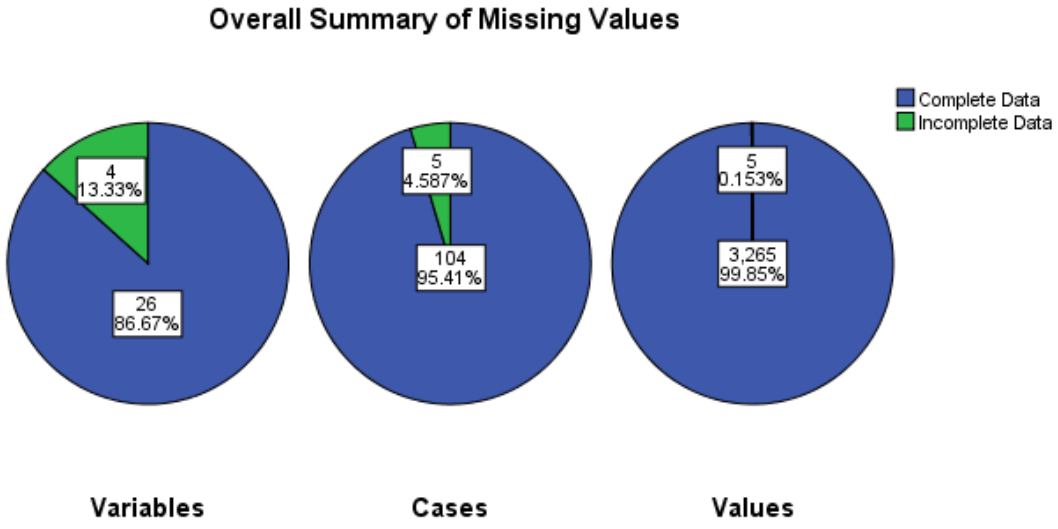


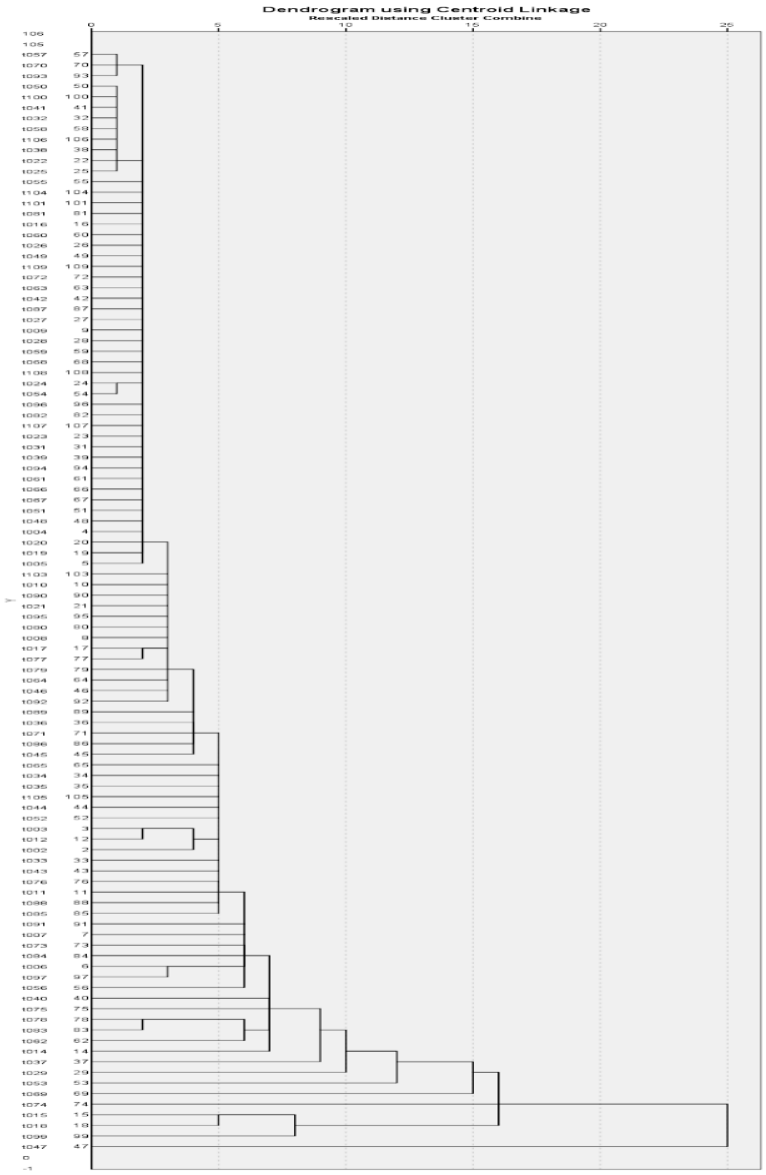
Table 59 Details of missing values

Variables	Missing Number	Missing Percentage
A4	2	1.8%
C6	1	.9%
B2	1	.9%
C1	1	.9%

Cluster Analysis for the Cartoon Ratings from Chinese Teachers. Cluster analysis A1 to E6 (30 variables). Cluster method: Centroid- square Euclidean distance. From *the graph* indicates that the teachers could be divided into two clusters or three clusters. The two clusters divide all the teachers into two groups based on their ratings. Cluster 1 includes all

the teachers except for teacher t047, who is in Cluster 2. This suggests that teacher t047 used the rating scale differently from the other teachers., the graph indicates that the teachers could be divided into two clusters or three clusters. The two clusters divide all the teachers into two groups based on their ratings. Cluster 1 includes all the teachers except for teacher t047, who is in Cluster 2. This suggests that teacher t047 used the rating scale differently from the other teachers.

Figure 39 Cluster analysis graph

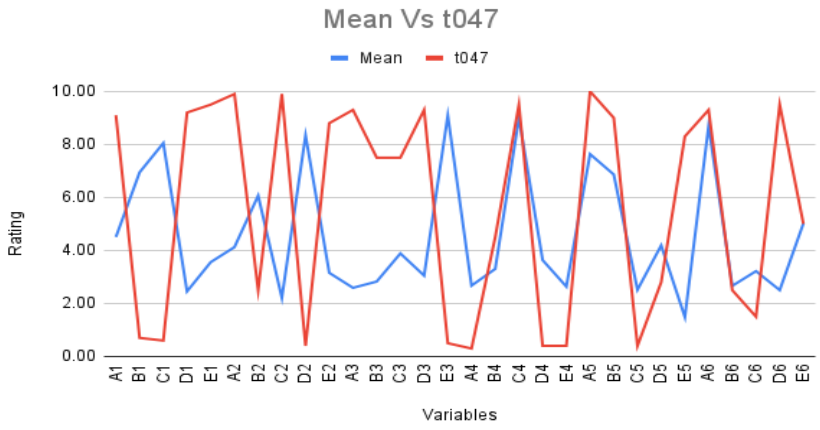


In the graphs below (*Figure40*), the horizontal axis represents the behaviours (variables) that teachers rated in the cartoons, and the vertical axis is the continuous rating scale from 0.00 to

10.00. The blue line represents the average ratings given by the other teachers to the behaviours in the cartoons, while the red line represents the ratings given by teacher t047 to the behaviours in the cartoons.

Figure 40 illustrates that t047 rated most of the cartoon behaviours differently compared to the rest of the sample. When the average ratings of variables (B1, C1, D2, E3, A4, D5, C6) are high, t047 rated them low. Conversely, when the average ratings of variables (D1, E1, A2, C2, E2, A3, B3, C3, D3, E5, D6, E6) are low, t047 rated them high. This discrepancy may indicate that either t047 has a very different perception of children's behaviours in the cartoons or t047 did not understand the rating scale. For instance, t047 might have rated positive behaviours as negative and negative behaviours as positive on the scale. Therefore, t047 is considered an outlier.

Figure 40 Teacher t047's ratings comparing with average ratings



The three clusters have identified three groups of raters: one group consists of t015 and t018, the second comprises t047, and the third includes the remaining teachers. This suggests that t015 and t018 rated the cartoon behaviours differently from t047 and the other teachers when compared to the average rating (mean). *Figure 41* and *Figure 42* reveal that both red lines (representing t015 and t018 ratings) are higher than the blue lines (average rating), indicating that these teachers assigned higher ratings to the behaviours in the cartoons than the rest of the

teachers. This could suggest that either these teachers are more accepting of the depicted behaviours, considering them mostly positive, or they may not have fully grasped the rating scale.

Teacher t015 assigned ratings ranging from 4.00 to 10.00, while teacher t018's ratings fell between 8.00 and 10.00. In contrast, the remaining teachers gave ratings ranging from 1.00 to 10.00. The Standard Deviation of the cartoon ratings from t015 and t018 is 0.05 and 0.14, respectively, which are the two lowest figures in the sample. This indicates that the thirty behaviours in the cartoons did not significantly differ in t015 and t018's ratings. Therefore, based on the analysis, the noisy data (t047, t015, t018) has been identified and excluded from the sample.

Figure 41 Teacher t015's ratings comparing with average ratings

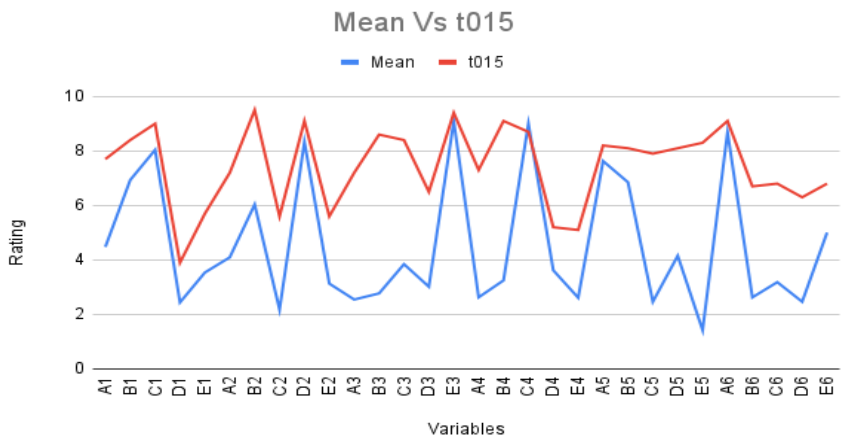
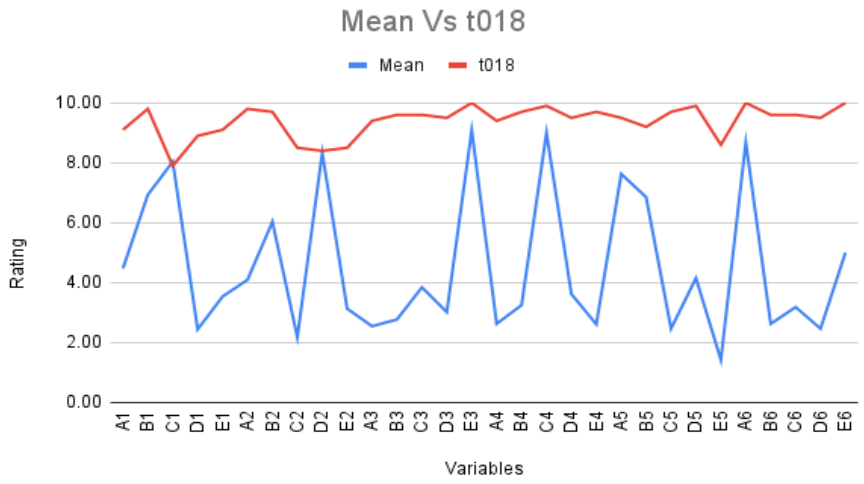


Figure 42 Teacher t018's ratings comparing with average ratings



9.2.2 Missing Data and Noisy Data in the English Sample for TRSC

In *Table 60*, three variables (E4, D4, B4) have 3 missing values in 1 case. *Figure 43* indicates that the percentage of missing data (variables 10%, cases 4.17%, values 0.42%) is very low. Therefore, the missing case with missing values does not significantly impact the power analysis, given that the rest of the data is available. The missing case is t05. These cases with missing data were retained for further analysis.

Figure 43 Overall summary of missing values

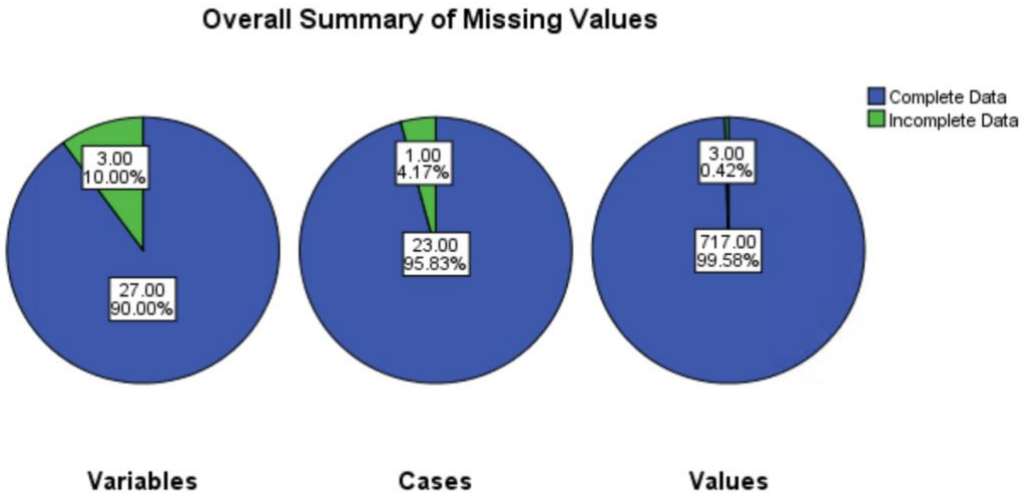


Table 60 Details of missing variables

Variables	Missing Number	Missing Percentage
E4	1	4.8%
D4	1	4.8%
B4	1	4.8%

Cluster Analysis for the Cartoon Ratings from English Teachers. Cluster analysis A1 to E6 (30 variables) graph is shown in *Figure 44*. Cluster method: Centroid- square Euclidean distance. From *Figure 44*, there appear to be two or three clusters worth investigating. The two clusters effectively divide all the teachers into two distinct groups based on their ratings. Cluster 1 comprises all the teachers except for teacher t24, who falls into cluster 2. This suggests that teacher t24 used the rating scale in a distinct manner compared to the other teachers. This distinction becomes apparent when comparing t24's ratings with the sample's average (mean) ratings.

In *Figure 45*, the horizontal axis represents the behaviours (variables) that teachers rated in the cartoons, while the vertical axis represents the continuous rating scale ranging from 0.00 to 10.00. The blue line represents the average ratings provided by the remaining teachers for the behaviours in the cartoons, while the red line represents the ratings given by teacher t24 for these behaviours.

Figure 44 Cluster analysis graph

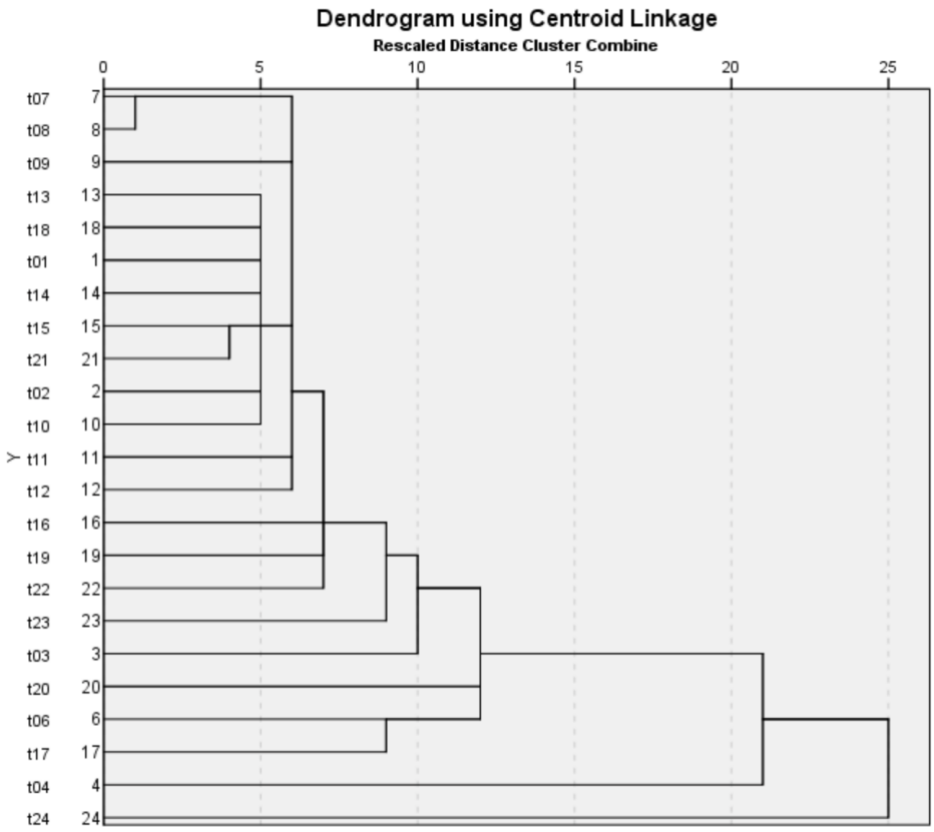


Figure 45 illustrates that t24 rated some of the cartoon behaviours differently compared to the rest of the sample. When examining the average ratings of variables (D2, C4, D4, E4, B6, E6), t024 rated them lower when these averages were higher. Conversely, when the average ratings of variables (A2, B2, A3, D3, C3, B6) were lower, t24 rated them higher. This discrepancy may indicate that either t24 has a very different perception of children's behaviours in the cartoons compared to other teachers, or t24 did not fully grasp the rating scale. For example, t24 might have rated positive behaviours negatively and negative behaviours positively on the scale. Therefore, t24 is considered an outlier and has been excluded from the data analysis.

The identification of three clusters has revealed three distinct groups of raters. One group includes t24, the other comprises t04, and the remainder forms the third group. This suggests that t04 rated the cartoon behaviours differently from t24 and the rest of the teachers. In *Figure 46*, it can be observed that most of t04's ratings (represented by the red line) align with the

average ratings (indicated by the blue line), except for a few behaviours' ratings (C6, D6). Therefore, t04 does not appear to be an obvious outlier and has been retained for further analysis.

Figure 45 Teacher t24's ratings comparing with average ratings

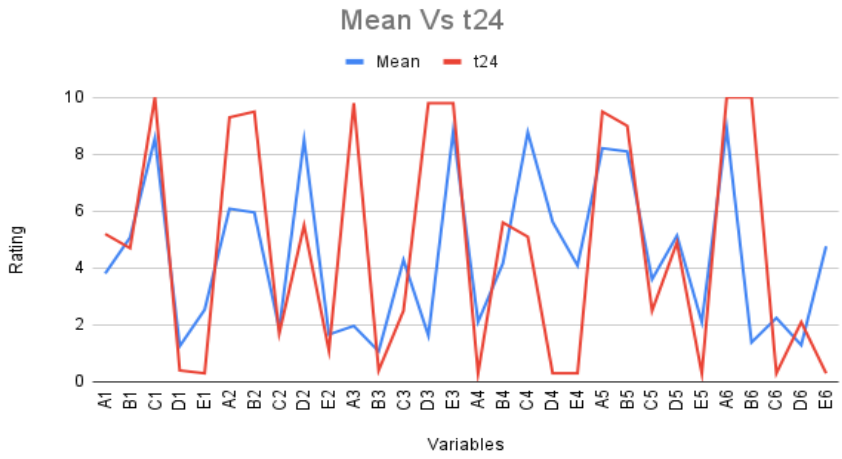
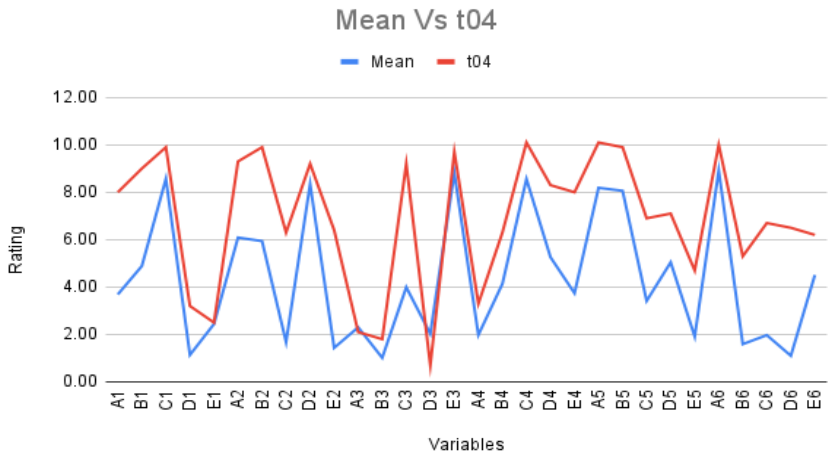


Figure 46 Teacher t04's ratings comparing with average ratings



9.3 Inter-rater Agreement (IRA) Analysis

To assess the extent of comparability in ratings of ADHD behaviours among teachers in China, Chinese teachers were requested to cross-rate children in their respective classes. There were 40 groups of children, each consisting of 10 children randomly selected from their respective classes. Each group of children received ratings from 2 to 5 teachers using The Teacher's

Rating Scale of Children's Behaviour. The study aimed to explore the degree of homogeneity or consensus among these ratings by calculating their inter-rater agreement.

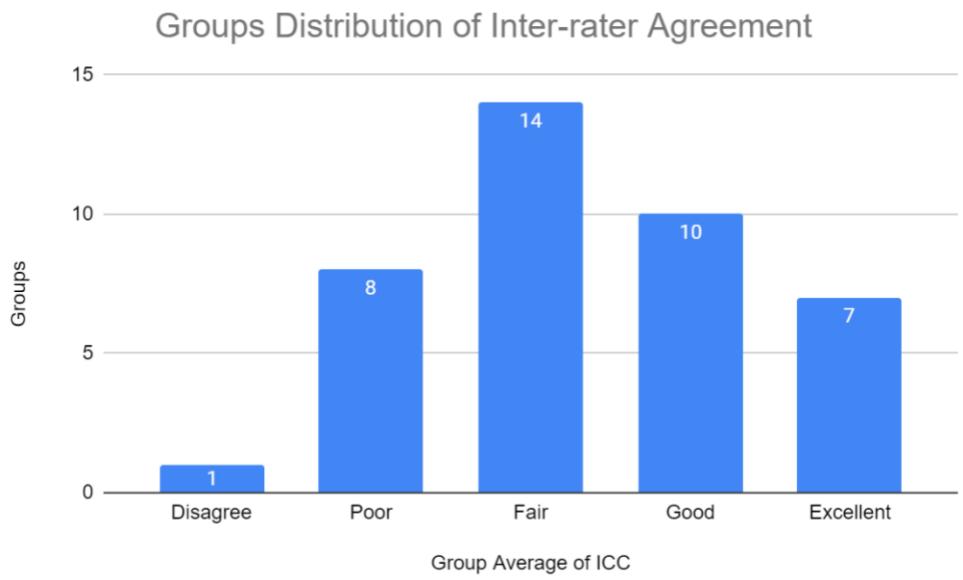
The inter-rater reliability of the 40 groups of teachers was evaluated using the intra-class correlation coefficient (ICC). ICC is a widely employed measure for assessing inter-rater agreement in ordinal, interval, and ratio variables (Hallgren, 2012). High ICC values signify a substantial level of agreement among raters, with the highest ICC value of 1 indicating perfect agreement, and a value of 0 suggesting random agreement. A negative ICC value suggests systematic disagreement among raters, and if three or more raters disagree, the ICC value will be less than -1 (Hallgren, 2012). Commonly cited benchmarks for ICC values categorize inter-rater agreement as poor for values less than .40, fair for values between .40 and .59, good for values between .60 and .74, and excellent for values between .75 and 1.0 (Cicchetti, 1994).

9.3.1 IRA Analysis of Each Group of Raters

The study examined the intra-class correlation coefficient (ICC) for each symptom and calculated the average ICC values for each group of raters. In this study, each group of teachers rated 10 randomly selected children from their respective classes, constituting a fully crossed rating design. The focus of the study was on teachers reaching consensus regarding the absolute values of their ratings, rather than seeking consistency with one another, where higher ratings from one teacher corresponded to higher ratings from another. The teachers were not chosen randomly, as each group consisted of teachers teaching the same class. The researcher aimed to understand the extent of agreement among teachers in the current study, rather than generalizing these ratings to a larger population of teachers. Given this research design and objective, the study conducted a two-way mixed model, assessing absolute agreement of average measures of ICC. Detailed results can be found in Appendix 11.

Figure 47 illustrates that one group of teachers (t052, t053) exhibited a negative ICC value, indicating systematic disagreement when rating the same group of children in their class. Additionally, 8 groups of teachers had ICC values less than .40, signifying poor agreement with their colleagues when rating the same group of children from their class. Furthermore, 14 groups of teachers had ICC values between 0.4 and .59, indicating a fair degree of agreement among their colleagues when rating the same group of children from their class. Another 10 groups of teachers had ICC values between .60 and .74, suggesting good agreement with their colleagues when rating the same group of children from their classes. Lastly, 7 groups of teachers had ICC values exceeding .75, demonstrating an excellent level of agreement with their colleagues when rating the same group of children from their class.

Figure 47 Groups distribution of Inter-rater agreement



9.3.2 IRA Analysis of Each Symptom

In *Figure 48*, five symptoms (Can't wait, Fidgety, Disorganised, Distracted, and Disruptive) exhibit ICC values ranging from .47 to .55, indicating that teachers have a fair level of agreement with their colleagues when rating the same group of children on these symptoms. For the symptom 'Inattentive,' the ICC value is .64, indicating that teachers have a good level

of agreement with their colleagues when rating the same group of children in their class on this symptom. *Table 61* presents the descriptive statistics for ICC values across the 40 groups of Chinese teachers, categorised by symptoms.

High agreement among teachers suggests that minimal measurement error has been introduced by independent raters, and the statistical power for subsequent analyses remains uncompromised. There are several reasons why teachers may exhibit low agreement when rating the same child for the same behaviours. One factor could be differing interpretations of the behaviours themselves, as in the case of the behaviour 'disorganised.' Some teachers may interpret a messy table as disorganised, while others may consider unpredictable agitation as disorganised.

Additionally, variations in the interpretation of rating categories can contribute to discrepancies. For instance, some teachers might regard 1 or 2 times a day as qualifying for 'Occasionally,' while others may categorise it as 'Very Often.' Individual teachers may also have varying levels of tolerance for specific behaviours. In this study, the symptom 'Inattentive' displayed the highest degree of agreement among the 40 groups of Chinese teachers, possibly indicating a substantial consensus among these teachers regarding children's inattentive behaviours and their tolerance levels. Conversely, the symptom 'Disorganised' exhibited the lowest degree of agreement among the same groups of Chinese teachers, suggesting that these teachers may interpret children's disorganised behaviours differently and have varying levels of tolerance for such behaviours.

Figure 48 Chinese teachers IRA of symptoms

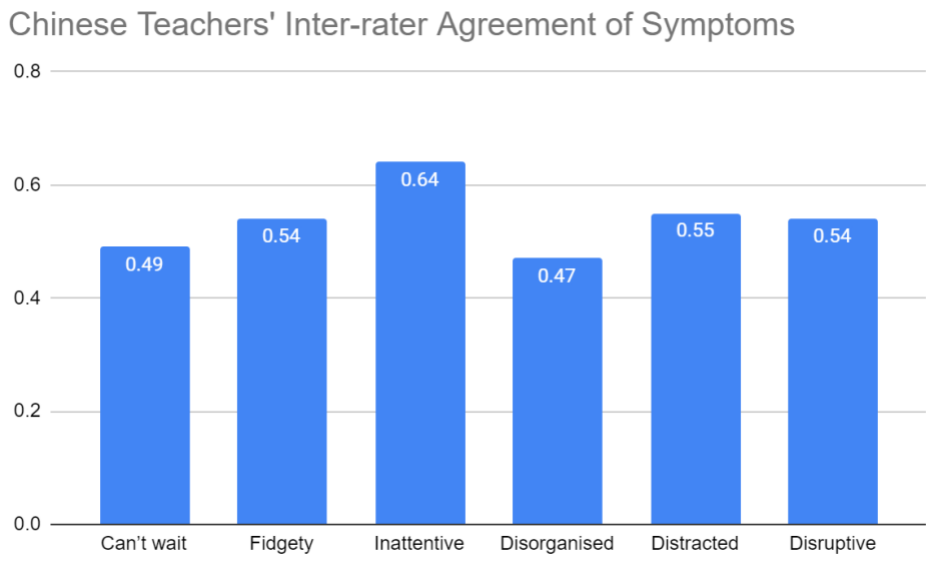


Table 61 Descriptive statistics for symptoms ICC across 40 groups of Chinese teachers

	Can't wait	Fidgety	Inattentive	Disorganised	Distracted	Disruptive
Min	-.92	-.41	-.41	.00	-.09	-.33
Max	.96	.94	.92	.89	.94	.94
SD	.40	.35	.27	.28	.28	.30
Mean	.49	.54	.64	.47	55	.54

9.4 Descriptive Analysis of Cartoon Behaviours and Children's ADHD Symptoms

In total, 133 teachers (109 from China and 24 from England) watched the cartoons and completed the surveys. Descriptive analysis of the ratings summarises the samples and measures. Furthermore, by comparing teachers' ratings of these cartoon characters' behaviours between the two countries, the researcher could identify cultural differences in the rating of children's ADHD behaviours between English and Chinese teachers. The surveys also collected teachers' background information, such as gender, teaching experience, and teaching position (class teacher or teaching assistant), to explore whether there are correlations between teachers' backgrounds and their ratings.

9.4.1 Descriptive Analysis of Cartoon Ratings

The analysis included 106 observations in the Chinese sample and 23 observations in the English sample. The mean and standard deviation for each behaviour is shown in *Table 62.* In the table, it's evident that the two groups of teachers rated 30 behaviours differently. However, to determine whether these differences are statistically significant and to understand the magnitude of these differences, further analysis is needed.

Table 62 The descriptive statistics of 30 animated behaviours that were rated by Chinese and English teachers

Cartoons	Behaviours		CN (N=106)		Eng (N=23)	
			Mean	SD	Mean	SD
Can't wait	A1	Leaving seat and asking for sweets	4.35	2.46	3.83	2.44
	B1	Shouting out for sweets	6.77	2.29	5.09	2.54
	C1	Waiting for	7.88	2.38	8.57	1.67
	D1	Leaving seat and eating sweets	2.33	2.15	1.26	1.48
	E1	Leaving seat and touching	3.42	2.41	2.55	1.84
Fidgety	A2	Moving body on the seat	3.96	2.26	6.09	2.16
	B2	Fiddling with a pencil and tapping feet	5.88	2.34	5.96	2.19
	C2	Moving chair	2.08	2.04	1.88	2.2
	D2	Sitting still	8.17	2.17	8.52	1.63
	E2	Fiddling with other's hair	3.02	2.42	1.66	1.77

Inattentive	A3	Doodling	2.43	2.01	1.98	1.69
	B3	Tearing paper	2.65	1.92	1.07	0.98
	C3	Daydreaming	3.72	2.13	4.28	2.48
	D3	Eating snack	2.9	2.35	1.64	2.14
	E3	Concentrating	8.9	1.24	8.83	1.14
Disorganized	A4	Late for school	2.51	2.19	2.11	2.52
	B4	Messy table	3.13	2.22	4.19	2.14
	C4	Tidy table	8.82	1.52	8.79	1.43
	D4	Looking for things	3.49	2.42	5.63	2.33
	E4	Nothing on the table	2.5	2.07	4.08	2.08
Distracted	A5	Having a look and back to the lesson	7.46	2.14	8.22	1.46
	B5	Looking at the butterfly all the time	6.7	2.38	8.12	1.37
	C5	Playing with the butterfly but staying on seat	2.35	2.13	3.63	2.91
	D5	Shouting out	4.02	2.59	5.15	2.27
	E5	Leaving seat to chaise the butterfly	1.33	1.88	2.12	2.18
Disruptive	A6	Concentrating on writing	8.51	1.74	8.93	1.15
	B6	Taking other’s pencil without asking	2.51	2.06	1.39	1.23

	C6	Tickling another child	3.06	2.1	2.25	2.16
	D6	Doodling on other's writing	2.35	1.84	1.28	1.5
	E6	Disturbing the teacher	4.86	3.04	4.76	2.7

9.4.2 Comparison of Cartoon Ratings between Chinese and English Teachers

To determine whether teachers from two significantly different cultural backgrounds rated ADHD behaviours differently and to identify which behaviours exhibited significant differences, further analysis is required. Independent Sample T-tests were conducted to compare the ratings of 30 variables (cartoon characters' behaviours) between Chinese and English teachers. By comparing the mean ratings between the two groups for each behaviour, any differences observed, along with a confidence interval of 95% CI ($p < 0.05$, two-tailed), indicate that the behaviour is significantly rated differently by Chinese and English teachers. The Independent Sample T-tests revealed that 13 behaviours exhibited significant differences in ratings between Chinese and English teachers, with p-values less than 0.05. The statistical details for these 13 behaviours are presented in *Table 63*.

Table 63 The statistics of 13 behaviours that were rated significantly differently by Chinese and English teachers.

Behaviour	Country	Mean	SD	t (df)	P<0.05 two tailed	Cohen's d
B1	CN	6.77	2.29	t (127) =3.13	.002**	.72
	ENG	5.09	2.54			
D1[a]	CN	2.33	2.15	t (45) =2.87	.006**	.52

	ENG	1.26	1.48			
A2	CN	3.96	2.26	t (127) =-4.12	<.001***	-.95
	ENG	6.09	2.16			
E2 [a]	CN	3.02	2.42	t (42) =3.12	.003**	.59
	ENG	1.66	1.77			
B3 [a]	CN	2.65	1.92	t (64) =5.72	<.001***	.88
	ENG	1.07	0.98			
D3	CN	2.9	2.35	t (127) =2.38	.02*	.55
	ENG	1.64	2.14			
B4	CN	3.13	2.22	t (126) =-2.05	.04*	-.48
	ENG	4.19	2.14			
D4	CN	3.49	2.42	t (126) =-3.79	<.001***	-.89
	ENG	5.63	2.33			
E4	CN	2.5	2.07	t (126) =-3.26	.001**	-.76
	ENG	4.08	2.08			
A5 [a]	CN	7.46	2.14	t (45) =-2.07	.04*	-.37
	ENG	8.22	1.46			
B5 [a]	CN	6.7	2.38	t (55) =-3.86	.007**	-.63

	ENG	8.12	1.37			
B6 [a]	CN	2.51	2.06	t (53) =3.45	.001**	.57
	ENG	1.39	1.23			
D6	CN	2.35	1.84	t (127) =2.60	.01*	.59
	ENG	1.28	1.5			

Note ($p<0.05$ *, $p<0.01$**, $p<0.001$***). [a] Welch test is reported because Levene's test indicated that the homogeneity of variances assumption was not met for this variable.

Figure 49 provides a visual representation of the ratings given by Chinese and English teachers for 13 specific behaviours, which exhibited significant differences between the two groups. The horizontal axis in the figure represents these 13 behaviours, with labels marked by ** signifying a p-value of less than 0.01 and labels marked by * indicating a p-value of less than 0.05. The red line corresponds to ratings provided by Chinese teachers, while the blue line represents ratings from English teachers.

Examining the chart, it becomes evident that for certain behaviours (B1, D1, E2, B3, D3, B6, D6), the red line is consistently higher than the blue line, indicating that Chinese teachers rated these behaviours significantly higher compared to their English counterparts. Conversely, for a different set of behaviours (A2, B4, D4, E4, A5, B5, C5), the blue line is consistently higher than the red line, implying that English teachers assigned significantly higher ratings for these behaviours than the Chinese teachers did.

In *Table 64,* the behaviours marked with a * are those Chinese teachers rated significantly higher than English teachers, while the remaining behaviours were rated significantly higher by English teachers compared to their Chinese counterparts. A closer examination reveals that

across three cartoons (Can't Wait, Inattentive, and Disruptive), Chinese teachers displayed greater acceptance for six specific behaviours, which included actions such as "Shouting out for sweets," "Eating snacks," and "Doodling on others' writing." In contrast, across two cartoons (Disorganised and Distracted), English teachers exhibited greater tolerance than their Chinese counterparts for six behaviours, which encompassed actions such as "Messy table" and "Looking at the butterfly all the time." It is worth noting that for the behaviour "Moving body on the seat" in the Fidgety cartoon, English teachers were more tolerant than Chinese teachers, whereas for the behaviour "Fiddling with others' hair" (E2), Chinese teachers displayed greater tolerance than their English counterparts. In summary, Chinese teachers found seven behaviours more acceptable than English teachers, while English teachers considered six behaviours more acceptable than their Chinese counterparts.

Figure 49 Thirteen behaviours that Chinese teachers and English teachers rated statistically different

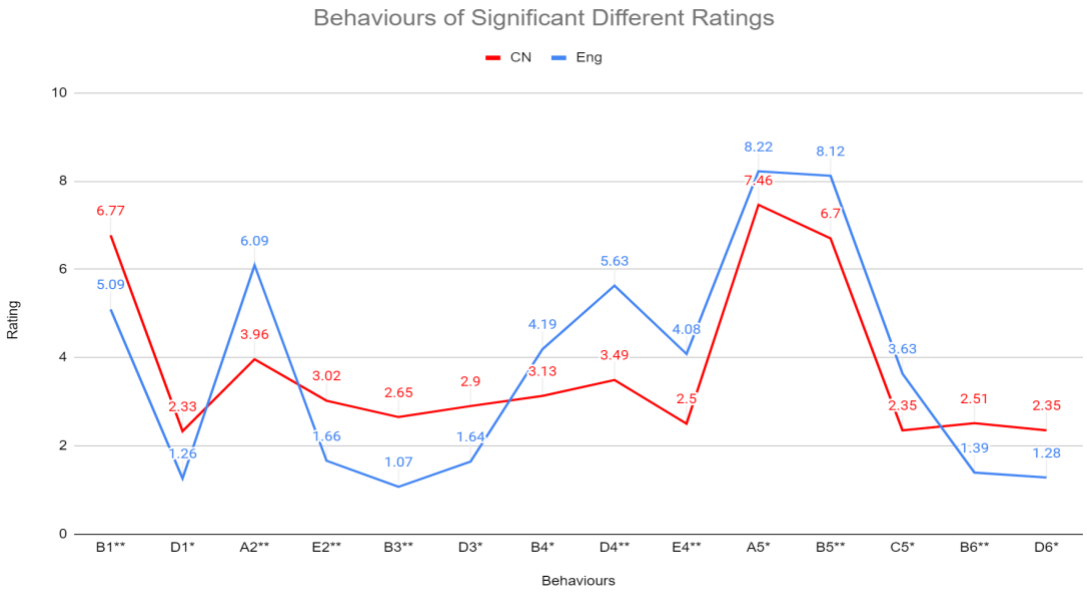


Table 64 Descriptions of 13 behaviours and rating differences across China and England

Cartoons	Behaviours	Rating Difference
Can’t Wait	B1* Shouting out for sweets	CN > ENG

	D1* Leaving seat and eating sweets	CN > ENG
Fidgety	A2 Moving body on the seat	CN < ENG
	E2* Fiddling with other's hair	CN > ENG
Inattentive	B3* Tearing paper	CN > ENG
	D3* Eating snack	CN > ENG
Disorganised	B4 Messy table	CN < ENG
	D4 Looking for things	CN < ENG
	E4 Nothing on table	CN < ENG
Distracted	A5 Having a looking and back to lesson	CN < ENG
	B5 Looking at the butterfly all the time	CN < ENG
Disruptive	B6* Taking other's pencil without asking	CN > ENG
	D6* Doodling on other's writing	CN > ENG

9.4.3 Correlations between Teachers' Background and Ratings

The study has identified 13 ADHD-type behaviours that were rated significantly differently by teachers from distinct cultural backgrounds. While cultural factors are acknowledged to influence teachers' perceptions of ADHD behaviours, it's essential to consider other potential variables affecting teachers' perceptions of children's behaviours, as elaborated upon in Chapter 3 Understanding ADHD Diagnosis and the Role of Culture and Raters. Consequently, the surveys gathered background information about the teachers, including their gender, teaching experience, position, and class size. The research employed two distinct statistical analyses to

explore potential correlations between teachers' backgrounds and their perceptions of children's behaviour.

To assess group differences, such as gender and teaching position, the study conducted Independent Sample T-tests to investigate whether significant variations existed in the ratings based on these categorical variables. For exploring correlations, like those related to teaching years and class size, Pearson's r was employed to determine whether any associations could be discerned between the ratings and these independent variables. In total, the analysis encompassed 106 Chinese teachers and 23 English teachers.

Gender and teachers' ratings

Examining the impact of gender on teachers' ratings, it is important to note that the Chinese sample comprised 95 female and 11 male teachers. The analysis indicated that there were no significant differences in their ratings. Conversely, the English sample comprised 20 female and 3 male teachers. However, due to the limited participation of male English teachers in the study, the sample size was deemed insufficient to draw robust conclusions. Consequently, we cannot definitively ascertain whether female teachers rated the animated ADHD behaviours significantly differently from their male counterparts.

Teaching Position and Teacher's Ratings

In the context of school settings, teachers hold diverse positions and responsibilities. Particularly in Chinese schools involved in this study, there is a notable variation in the number of Year 2 classes, with some institutions having as many as 11 of them. This often results in a collaborative teaching approach, where Year 2 education is a team effort involving class teachers and subject teachers. In this setup, each teacher is assigned responsibility for instructing a specific subject from the national curriculum. However, class teachers carry additional duties beyond subject instruction. They manage various administrative tasks, handle

daily assemblies, prepare teacher reports, and engage in parent meetings, all focused on a single Year 2 class. Conversely, subject teachers are primarily tasked with subject-specific teaching across multiple Year 2 classes.

In contrast, the role of class teachers in England encompasses a wider array of responsibilities. They not only deliver instruction across all Year 2 subjects but also oversee various school activities. Depending on the specific class, some Year 2 classes in England may also have teaching assistants to support the class teacher in their teaching duties. Given these varying roles and workloads within schools, it's intriguing to explore whether these factors influence how teachers perceive ADHD behaviours in children. To delve into this matter, we conducted a One Way ANOVA to compare teachers' ratings.

Within the Chinese sample, there were 58 class teachers and 48 subject teachers. Interestingly, a significant difference ($p = 0.04$, $p < 0.05$) emerged between the ratings provided by Chinese class teachers (Mean = 8.84, SD = 1.12) and those offered by Chinese subject teachers (Mean = 8.10, SD = 2.22) in relation to behaviour A6, specifically 'Concentrating on writing.' Here, Chinese class teachers rated this behaviour significantly higher than Chinese subject teachers. However, it's essential to consider the broader context, as we compared ratings for 30 distinct behaviours. With a significance level set at 0.05, it's reasonable to expect that one or two significant differences may arise by chance. Consequently, the observed variance in behaviour A6 may not necessarily be statistically significant and could potentially be attributed to chance. Conversely, in the English sample, which included 15 class teachers and 8 teaching assistants, no significant difference emerged between the ratings provided by class teachers and teaching assistants across the 30 behaviours under consideration. Therefore, this study did not find relation between teacher's teaching position and their ratings to animated children’s behaviours.

Teaching Experience and Teachers' Ratings

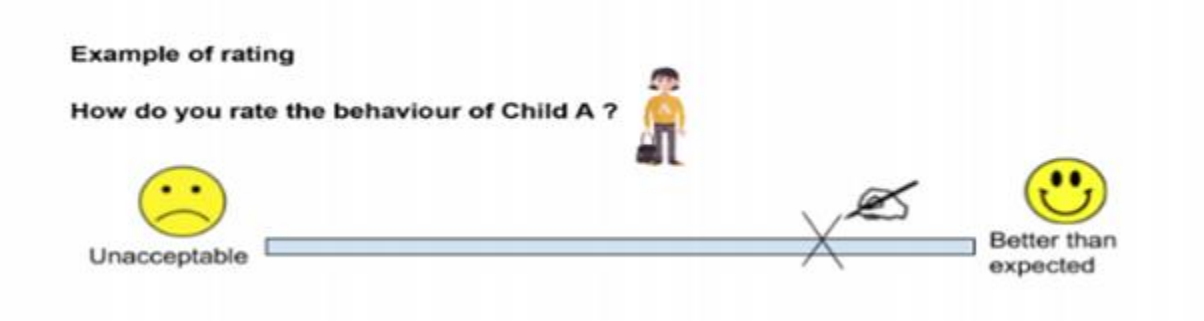
As discussed in chapter 3, the study explored how teaching experience might influence teachers' perceptions of children's problematic behaviours. Different levels of teaching experience can correspond to varying degrees of tolerance towards different types of problematic behaviours (Borg, 1998; Kokkinos et al., 2004). Within this sample, Chinese teachers exhibited a range of teaching experience, spanning from 36 years to 1 year, with an average of 16 years. English teachers, on the other hand, showcased teaching experience that varied from 20 years to 1 year, with an average of 9 years. To delve into these dynamics, we conducted a Pearson correlation analysis within SPSS to uncover potential links between teachers' years of teaching experience and their evaluations of the 30 animated children's ADHD behaviours.

As depicted on the semantic differential scale (*Figure 50*), higher behaviour ratings signify greater acceptability of that behaviour by the teacher. Hence, a significant positive correlation between the two variables (teaching years and behaviour rating) would imply that teachers with more extensive teaching experience tend to view these behaviours as more acceptable. Conversely, a significant negative correlation between the two variables (teaching years and behaviour rating) would suggest that teachers with greater teaching experience are more inclined to perceive these behaviours in a stricter light. It's worth noting that this could also be influenced by the teacher's age, with older teachers potentially holding more traditional views regarding children's behaviours.

For interpreting the correlation coefficients, I have adopted a conventional approach, which classifies correlations as follows: .00-.10 as negligible correlation, .10-.39 as a weak

correlation, .40-.69 as a moderate correlation, .70-.89 as a strong correlation, and .90-1.00 as a very strong correlation (Schober et al., 2018).

Figure 50 The semantic differential scale for animated children's ADHD behaviours



There are statistically significant correlations between the Chinese teacher's teaching year and their ratings of behaviours A1, E2, A3, and D4. The statistical data is shown in Table *65*. The Chinese teachers' ratings have relatively weak negative correlations with the behaviours in the table, which indicates that with more teaching experiences, Chinese teachers tend to rate those behaviours more harshly. The full correlations see Appendix 20.

Table 65 The correlation between Teaching Years and Chinese teachers' Ratings

	Behaviour	r	P (0.05, two-tailed)	n	Mean
A1	Leaving seat and asking for sweets	-.21*	.03	106	4.35
E2	Fiddling with other's hair	-.23*	.02	106	3.02
A3	Doodling	-.28**	.00	106	2.43
D4	Looking for things	-.25*	.01	106	3.49

There are statistically significant correlations between the English teacher's teaching years and their ratings of the behaviours A2, E3, C4, D5, and A6. The statistical data is shown in Table

66. There are moderate negative correlations between the English teacher's ratings of these behaviours with the teacher's teaching years. The negative correlations indicate that with more teaching experiences, English teachers tend to rate those behaviours more harshly. The full correlations see Appendix 18.

Table 66 The correlation between Teaching Years and English teachers' Ratings

	Behaviour	r	P (0.05, two-tailed)	n	Mean
A2	Moving body on seat	-.53*	.01	23	6.09
E3	Concentrating	-.48*	.02	23	8.83
C4	Tidy table	-.46*	.03	23	8.79
D5	Shouting out	-.56**	.01	23	5.15
A6	Concentrating on writing	-.46*	.03	23	8.93

In conclusion, the analysis reveals significant correlations between teaching experience and teacher ratings for specific ADHD-type behaviours among both Chinese and English teachers. Interestingly, this study indicates that experienced teachers tend to adopt a more critical stance towards children's behaviours in both countries.

Class Size and Teachers' Ratings

Studies indicate that class size affects children's learning and teaching practices (e.g., Blatchford et al., 2001; Blatchford et al., 2002; Blatchford, 2003; David Pedder, 2008). Different class sizes lead to variations in teaching practices, workloads, and teaching-related stress among educators. Moreover, class size can also influence teachers' perceptions of children's behaviours (Meixia Ding, Yeping Li, Xiaobao Li & Gerald Kulm, 2008).

In this sample, class sizes in Chinese schools vary from 32 to 60, with an average size of 47, while in English-participating schools, class sizes range from 10 to 30, with an average size of 23. To investigate whether there is a correlation between class size and teachers' ratings, we conducted a Pearson correlation analysis between class size and the 30 animated ADHD behaviours that teachers rated. A positive correlation between the two variables (class size and teachers' rating) suggests that larger class sizes are associated with greater acceptance of these behaviours by teachers. Conversely, a negative correlation between the two variables indicates that larger class sizes are linked to less tolerance of these behaviours.

In *Table 67*, the analysis reveals a statistically significant correlation between class size and Chinese teachers' ratings of behaviour B2 (fiddling with a pencil and tapping feet). However, this correlation is relatively weak and negative. This suggests that Chinese teachers in relatively large classes tend to be less tolerant of the behaviour of fiddling with a pencil and tapping feet in the classroom. It's worth noting that there are 30 correlations calculated between behavioural ratings and class size. Given a significance level of p = 0.05, we would expect one or two significant correlations to occur by chance. Consequently, the observed correlation may not be very convincing when considering the likelihood of it occurring by chance alone. For the complete set of correlations, please refer to Appendix 21.

Table 67 The correlation between Class Size and Chinese teachers' Ratings

	Behaviour	r	P (0.05, 2-tailed)	n	Mean
B2	Fiddling with a pencil and tapping feet	-.24*	.01	106	5.88

In *Table 68*, the analysis reveals a statistically significant correlation between class size and the ratings given by English teachers. The two variables exhibit a moderate positive correlation ($r=0.49$, $p<0.05$). This indicates that English teachers who instruct relatively large classes tend to be more accepting of children's behaviour, specifically, shouting out. When compared to teaching smaller classes, educators dealing with larger class sizes often face increased workloads and heightened stress levels. In larger classes, it becomes more challenging for a teacher to closely manage every interaction. Consequently, they may exhibit greater tolerance towards instances of students shouting out in larger classes than in smaller ones. For the complete set of correlations, please refer to Appendix 19.

Table 68 The correlation between Class Size and English teachers' Ratings

	Behaviour	r	P (0.05, two-tailed)	n	Mean
D5	Shouting out	.49*	.02	23	5.15

9.4.4 Descriptive Analysis of Children's ADHD Symptoms

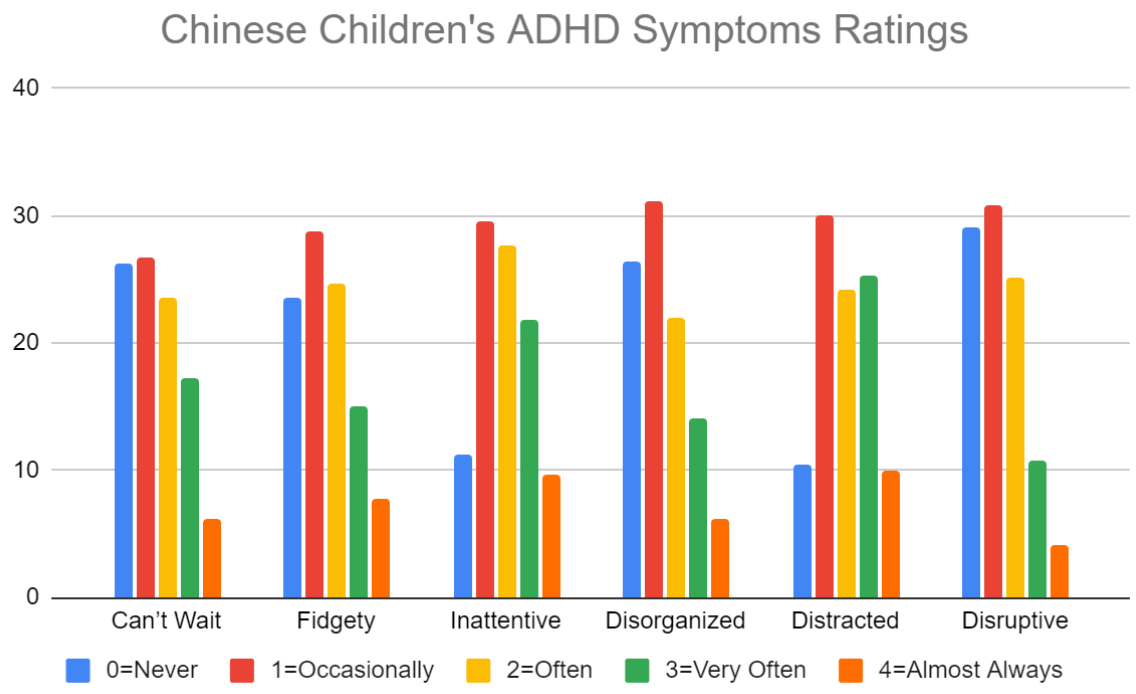
Chinese Children's ADHD Symptoms Ratings. In the Chinese sample, 550 children participated, and 400 were cross rated. When analysed teachers' ratings of their children's behaviour, the study selected teachers' ratings of cross-rated children based on no missing ratings and fit for the Rasch Model. In total, the study had 550 children in analysis. The English participating children were not cross rated in this research. In total, excluding the noisy data and unfit teachers from Rasch analysis, 210 children were rated by their participating teachers. The descriptive analysis is shown in *Table 69* and *Figure 51* below.

Table 69 Chinese children's behaviours ratings frequencies %

Symptoms	0=Never	1=Occasionally	2=Often	3=Very Often	4=Almost Always

Can't Wait	26.2	26.7	23.6	17.3	6.2
Fidgety	23.5	28.7	24.7	15.1	7.8
Inattentive	11.3	29.5	27.6	21.8	9.6
Disorganized	26.4	31.1	22.0	14.0	6.2
Distracted	10.5	30.0	24.2	25.3	10.0
Disruptive	29.1	30.9	25.1	10.7	4.2

Figure 51 Chinese children's ADHD symptoms ratings



Research shows that boys are three times more likely to be diagnosed with ADHD than girls (APA, 2000). In the Chinese sample, there are 294 boys and 256 girls. From *Table 70* and, *Figure 52*, *Figure 53*, we can see that more percentage of boys received rank 2 (Often), rank 3 (Very Often) and 4 (Almost Always) of the six symptoms than the percentage of girls. However, more percentage of girls received rank 0 (Never) and rank 1 (Occasionally) of the symptoms

than the percentage of boys. To investigate this gender difference, the study conducted an Independent Sample T-test to compare the two gender groups of children's behaviour ratings.

Table 70 Chinese boys and girls rating frequencies%

Symptoms	0=Never	1=Occasionally	2=Often	3=Very Often	4=Almost Always
Can't Wait	18.0 / 35.5	27.6 / 25.8	24.8 / 22.3	20.7 / 13.3	8.8 / 3.1
Fidgety	17.0 / 30.9	25.9 / 32.0	26.2 / 23.0	20.1 / 9.4	10.9 / 4.3
Inattentive	7.8 / 15.2	25.2 / 34.4	29.6 / 25.4	25.2 / 18.0	12.2 / 6.6
Disorganized	19.4 / 34.4	29.3 / 33.2	23.5 / 20.3	18.0 / 9.4	9.9 / 2.0
Distracted	7.5 / 14.1	24.5 / 36.3	26.2 / 21.9	28.2 / 21.9	13.6 / 5.9
Disruptive	20.4 / 39.1	28.6 / 33.6	31.0 / 18.4	14.3 / 6.6	5.8 / 2.3

Figure 52 Chinese boys ADHD symptoms ratings

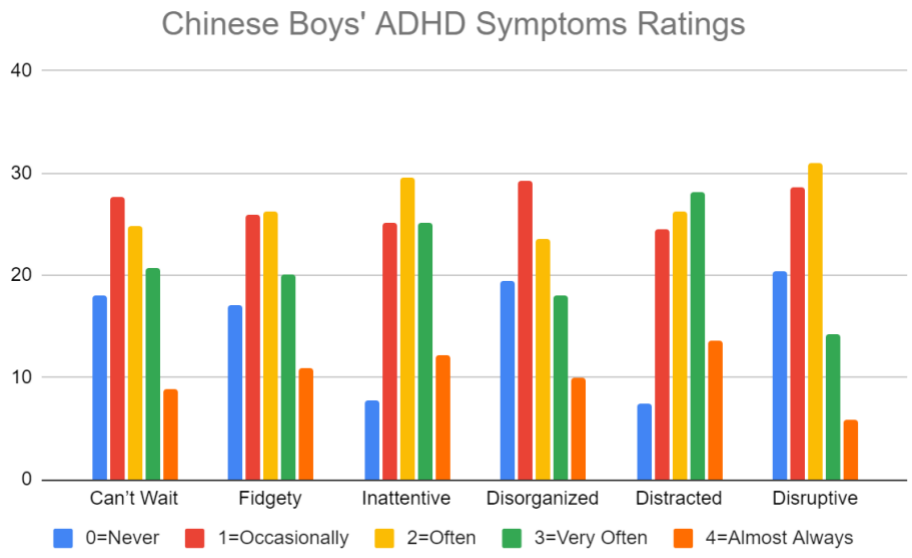
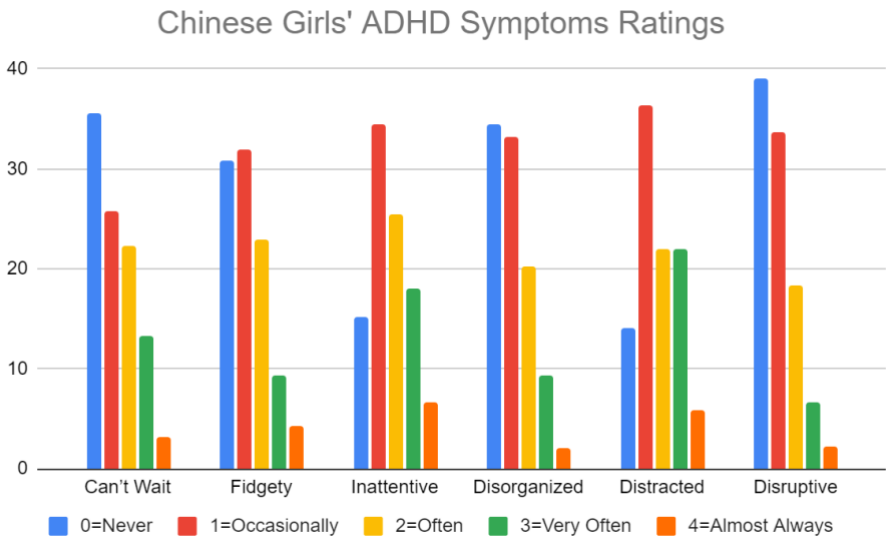


Figure 53 Chinese girls' ADHD symptoms ratings



The analysis shows significant differences between boys and girls across all six symptoms (Can't Wait, Fidgety, Inattentive, Disorganised, Distracted, Disruptive) and total score. Boys' ratings of the six symptoms and total scores are significantly higher than girls' ratings of the symptoms and total scores. The results are shown in *Figure 54* and *Table 71* below.

Figure 54 The average score of Chinese boys and girls

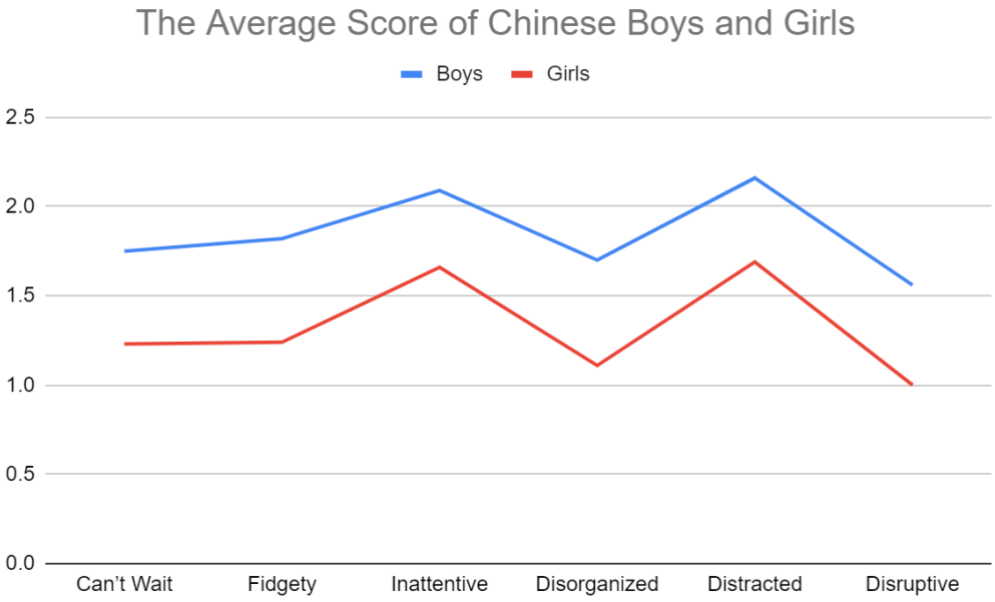


Table 71 The differences between Chinese boys and girls

	Boys		Girls					
	M	SD	M	SD	df	t	P 2-tailed	Cohen's d
Can't Wait	1.75	1.23	1.23	1.16	548	-5.11	<.001***	-.44
Fidgety[a]	1.82	1.24	1.24	1.12	546	-5.75	<.001***	-.49
Inattentive	2.09	1.14	1.66	1.14	547	-4.36	<.001***	-.37
Disorganized[a]	1.70	1.25	1.11	1.05	546	-6.02	<.001***	-.51
Distracted	2.16	1.16	1.69	1.14	548	-4.78	<.001***	-.41
Disruptive[a]	1.56	1.14	1.00	1.03	547	-6.16	<.001***	-.52
Total Score[a]	11.8	6.17	7.90	5.52	548	-6.37	<.001***	-.54

Note. [a] Welch test is reported because Levene's test indicated that the homogeneity of variances assumption was not met for this variable.

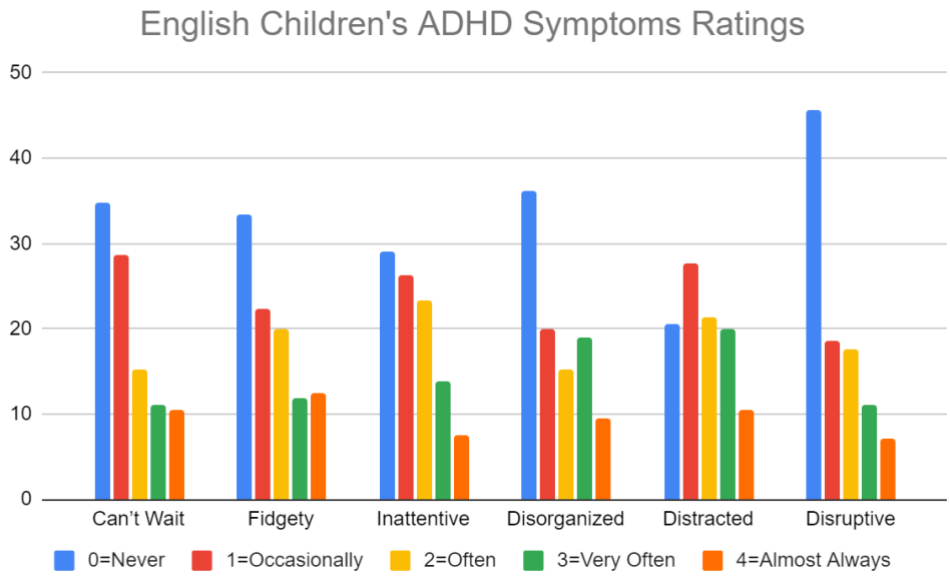
English Children's ADHD Symptoms Ratings. In the English sample, excluding ratings of unfit teachers for Rasch analysis, there were 210 children in the analysis. The English children's ADHD symptoms rating frequencies is shown in *Table 72* and *Figure 55*.

Table 72 English children's ADHD symptoms ratings frequencies%

Symptoms	0=Never	1=Occasionally	2=Often	3=Very Often	4=Almost Always
Can't Wait	34.8	28.6	15.2	11.0	10.5
Fidgety	33.3	22.4	20.0	11.9	12.4
Inattentive	29.0	26.2	23.3	13.8	7.6
Disorganized	36.2	20.0	15.2	19.0	9.5
Distracted	20.5	27.6	21.4	20.0	10.5

Disruptive	45.7	18.6	17.6	11.0	7.1

Figure 55 English children's ADHD symptoms ratings



There were 121 boys and 109 girls whom their teachers rated. In *Table 73*, *Figure 56* and *Figure 57*, we can see that, Similar to the Chinese sample, more percentage of boys received rank 2 (Often), rank 3 (Very Often) and 4 (Almost Always) of the six symptoms than the percentage of girls. However, more percentage of girls received rank 0 (Never) and rank 1 (Occasionally) of the symptoms than the percentage of boys. To investigate this gender difference, we conducted an Independent Sample T-test to compare the two gender groups of children's ratings.

Table 73 English boys' and girls' ratings frequencies

Symptoms	0=Never	1=Occasionally	2=Often	3=Very Often	4=Almost Always
Can't Wait	29.7 / 40.4	27.9 / 29.3	16.2 / 14.1	12.6 / 9.1	13.5 / 7.1
Fidgety	23.4 / 44.4	19.8 / 25.3	21.6 / 18.2	16.2 / 7.1	18.9 / 5.1

Inattentive	21.6 / 37.4	26.1 / 26.3	23.4 / 23.2	18.0 / 9.1	10.8 / 4.0
Disorganized	29.7 / 43.4	15.3 / 25.3	18.0 / 12.1	24.3 / 13.1	12.6 / 6.1
Distracted	12.6 / 29.3	26.1 / 29.3	23.4 / 19.2	24.3 / 15.2	13.5 / 7.1
Disruptive	33.3 / 59.6	18.0 / 19.2	26.1 / 8.1	12.6 / 9.1	9.9 / 4.0

Figure 56 English boys' ADHD symptoms ratings

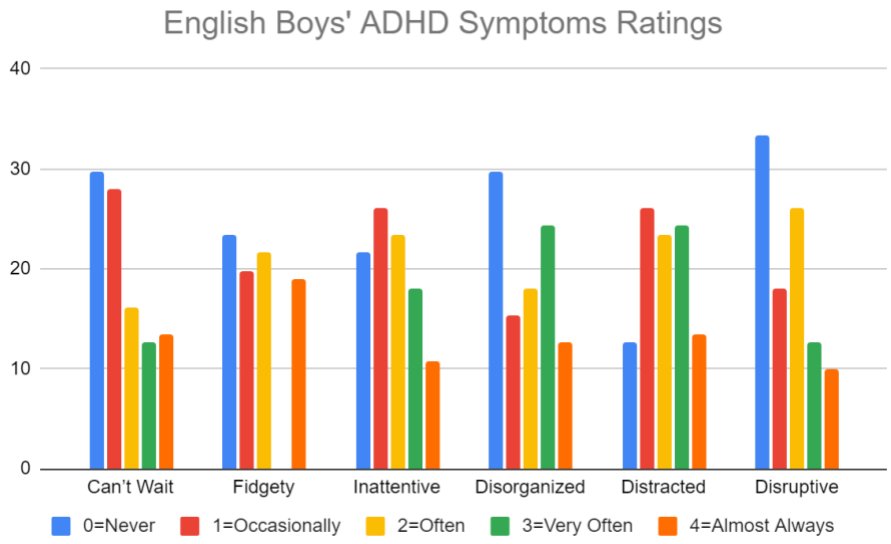
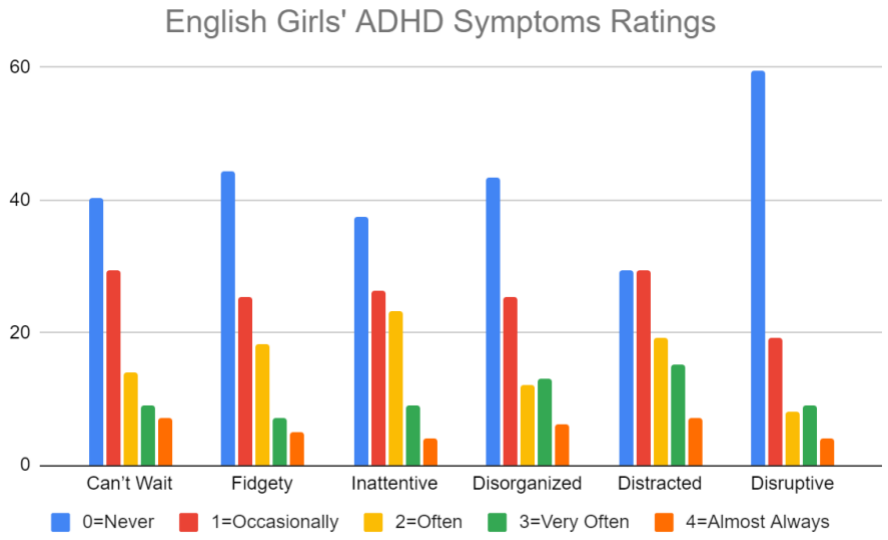


Figure 57 English girls ADHD symptoms ratings



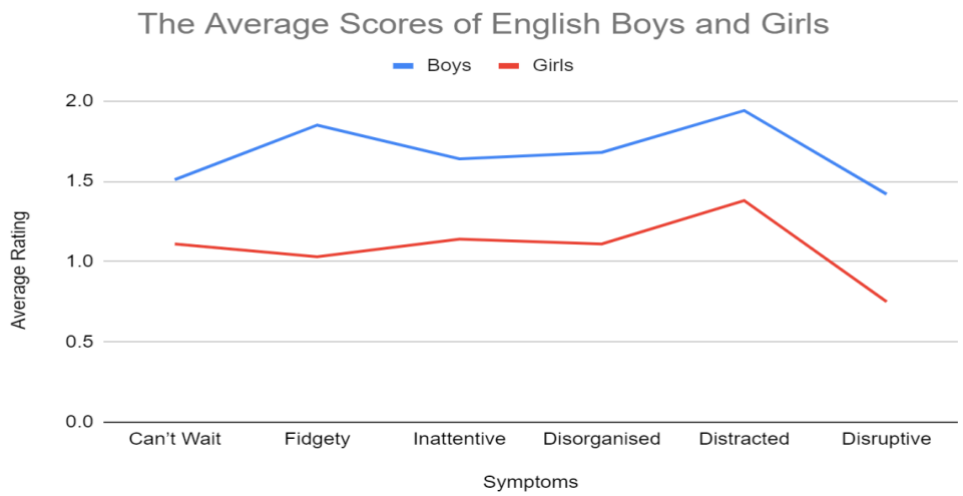
The analysis shows statistically significant differences between boys and girls across all six symptoms (Can't Wait, Fidgety, Inattentive, Disorganised, Distracted, Disruptive) and total score. Boys' ratings of the six symptoms and total scores are significantly higher than girls' ratings of the symptoms and total scores. The results are shown in *Table 74* and *Figure 58* below.

Table 74 The differences between English boys and girls

	Boys		Girls					
	M	SD	M	SD	df	t	P 2-tailed	Cohen's d
Can't Wait[a]	1.52	1.39	1.13	1.24	208	-2.16	.03*	-.30
Fidgety[a]	1.87	1.43	1.03	1.17	207	-4.68	<.001***	-.64
Inattentive	1.70	1.29	1.16	1.15	208	-3.19	<.01**	-.44
Disorganized[a]	1.75	1.43	1.13	1.28	208	-3.30	<.01**	-.45
Distracted	2.00	1.25	1.41	1.25	208	-3.39	<.01**	-.47
Disruptive[a]	1.48	1.33	.79	1.17	208	-3.99	<.001***	-.55
Total Score	10.32	6.80	6.66	6.15	208	-4.08	<.001***	-.56

Note. [a] Welch test is reported because Levene's test indicated that the homogeneity of variances assumption was not met for this variable.

Figure 58 The average score of English boys and girls



9.4.5 Comparison of Children's Symptoms between Chinese and English Children

The Chinese and English teachers' ratings of their children were compared to investigate whether there were significant differences across the six symptoms and total scores between the two culturally different groups of children. The study conducted Independent T-tests, and the results are in *Table 75*. The study found that there is a significant difference between the ratings of Chinese and English children's behaviours on symptom Inattentive, t (354) =4.45, p<0.001, d=.37. The effect size for this analysis (d=.37) was found in the range of a medium effect (Cohen, 1988). The results indicate that the ratings of Chinese children's inattentive behaviour (M=1.79, SD=1.23) are significantly higher than the ratings of English children's inattentive behaviour (M=1.40, SD=1.26).

Meanwhile, on symptom 'Distracted', there is a significant difference between the ratings of Chinese and English children's behaviours, t (350) =2.14, p=.03, d=.18. The effect size for this analysis (d=.18) was found to be small (Cohen, 1988). The results indicate that Chinese children (M=1.94, SD=1.17) were rated significantly higher than English children (M=1.72, SD=1.28) on the symptom 'Distracted'. However, there is no significant difference in the total score across the two groups of children. A two-way ANOVA analysis was carried out to investigate the effect of gender and country on children's total scores. It revealed that gender

significantly affects children's ratings (p<.001). Boys' total scores are significantly higher than girls' total scores in *Table 76.* However, there is no significant interaction effect between gender and country (p=.62). As we can see in *Figure 59*, the green line (boys' total score) is higher than the blue line (girls' total score) across both countries.

Table 75 The differences between Chinese and English children's ratings

	CN		EN					
	M	SD	M	SD	df	t	P 2-tailed	Cohen's d
Can't Wait	1.51	1.22	1.34	1.33	758	1.65	.10	.13
Fidgety[a]	1.55	1.22	1.48	1.38	341	.68	.50	.06
Inattentive[a]	1.89	1.16	1.45	1.25	354	4.45	<.001***	.37
Disorganized[a]	1.42	1.20	1.46	1.39	334	-.31	.76	-.03
Distracted[a]	1.94	1.17	1.72	1.28	350	2.14	.03*	.18
Disruptive[a]	1.30	1.12	1.15	1.30	334	1.45	.15	.13
Total Score[a]	9.60	6.08	8.60	6.74	345	1.89	.06	.16

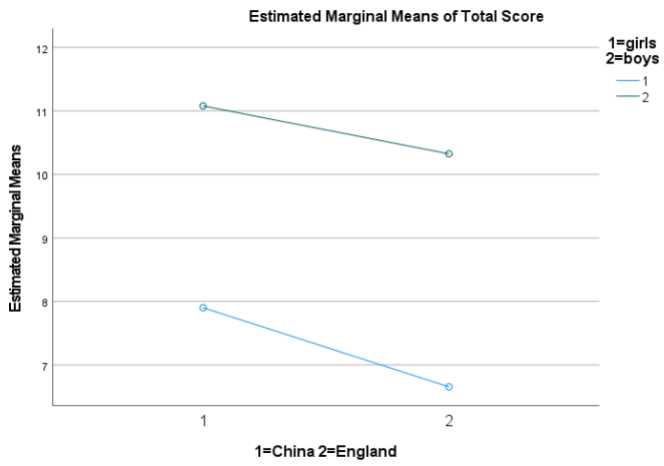
Note. [a] Welch test is reported because Levene's test indicated that the homogeneity of variances assumption was not met for this variable.

Table 76 Descriptive analysis of gender differences across China and England.

Gender	Country	Mean	SD
Girls	CN	7.90	.38
	EN	6.66	.61
Boys	CN	11.08	.35
	EN	10.32	.58

Note. CN=China, EN=England

Figure 59 Gender and country factors on children's total score



9.5 Measuring Teachers' Leniency across China and England

The study investigated the reliability of the Teacher's Rating Scale of Cartoons for measuring the Chinese and English samples. It has proved that this scale measures what it intends to measure. As the study aimed to get the measurements of Chinese and English teachers' leniency towards ADHD behaviours on one logit scale, the study combined Chinese and English data to get the measurements. The combined Chinese and English data include 129 persons (106 Chinese, 23 English), and the scale includes 30 items. Before conducting the DIF analysis with Winsteps software, 23 misfit persons (21 Chinese, 2 English) were excluded from the analysis. These misfit persons' responses do not fit the Rasch model, and their estimates have big errors. Furthermore, these misfit persons' responses produce noises to generate accurate measures. Therefore, the analysis excluded them from further analysis.

9.5.1 Calibrating the Instrument for Measuring Both Countries on One Scale

However, an essential aspect of test function is to provide evidence that a test is fair or free of bias (Bachman, 2005). Many factors could systematically cause a test to favour one group against another group of test takers. In many cases, external factors such as age, gender, academic background, and ethnicity influence people's performance in a test rather than their abilities (Ownby & Waldrop-Valverde, 2013; Pae, 2004; Gnaldi& Bacci, 2016;). From the previous analysis of the difference in raw data between Chinese and English teachers, some

behaviours teachers have rated significantly different between the two countries. These items (behaviours) could potentially have different difficulty estimates across the two groups of teachers. DIF analysis will detect the items functioning differently across Chinese and English teachers. The analysis needs to exclude DIF items from the measurement to get fair measures across the two groups of teachers.

By re-estimating item difficulty, while holding everything else constant, Winsteps estimates the new logit values for each subgroup (Chinese and English) using a logit difference method (Linacre, 2018). *Table 77* shows DIF MEASURE's new estimated item difficulty for each group. DIF CONTRAST is the difference in item difficulty between the two subgroups. Positive DIF Contrast indicates that the item is more difficult for Chinese teachers to endorse, and negative DIF CONTRAST indicates that the item is easier for Chinese teachers. Whether the difference between the DIF logit is statistically significant is determined by the Rasch-Welsh t statistics. The study adopted category C as our cut-off point for DIF items as a survey. Therefore, DIF analysis identified items with DIF contrast size equal to or bigger than .64 logit, and the Rasch-Welch t is significant ($p<.05$). *Figure 60* shows the items in category C with DIF contrast size and statistical significance.

The results of the DIF analysis show that there were 10 items flagged for DIF. In *Table 77*, the top two items, B1 and E2, are more difficult for the English teachers (negative DIF contrast with DIF contrast size range of -1.10 to -1.91). English teachers rated them significantly lower than Chinese teachers. It indicates that English teachers are less tolerant of these behaviours than Chinese teachers. While the rest 8 items are more difficult for the Chinese teachers (positive DIF contrast with DIF contrast size range of .85 to 1.97). Chinese teachers rated them significantly lower than English teachers. It indicates that Chinese teachers are less tolerant of

these behaviours than English teachers. As these 10 items function differently across two groups of teachers, they are evidence of threatening the invariance of measurement and could cause potential bias. Therefore, the analysis has excluded these 10 items from the measurement to have a fair measurement for each subgroup. The final Teacher's Rating Scale of Cartoons includes 20 items.

Table 77 DIF statistics for country

DIF Measure (logit)		DIF Contrast	Prob	Items	Behaviour's Description
CN	EN				
-1.35	-.25	-1.10	.01	B1	Shouting out for sweets
1.13	3.04	-1.91	.02	E2	Fiddling with other's hair
.46	-.95	1.41	.00	A2	Moving body on seat
.81	-.83	1.64	.00	D4	Looking for things
1.84	.43	1.41	.00	E4	Nothing on the table
-1.87	-3.83	1.97	.02	B5	Looking at the butterfly all the time
1.68	.67	1.01	.02	C5	Playing with the butterfly but staying on seat
.49	-.37	.86	.03	D5	Shouting out
.95	-.13	1.08	.01	C3	Daydreaming
1.20	-.02	1.22	.00	B4	Messy table

Figure 60 The interpretation of DIF contrast and statistical significance

ETS DIF Category	with DIF Contrast and DIF Statistical Significance	
C = moderate to large	|DIF| ≥ 0.64 logits	prob(|DIF| ≤ 0.43 logits) ≤ .05 (2-sided) *approximately: |DIF| > 0.43 logits + 2 * DIF S.E.*
B = slight to moderate	|DIF| ≥ 0.43 logits	prob(|DIF| = 0 logits) ≤ .05 (2-sided) *approximately: |DIF| > 2 * DIF S.E*
A = negligible	-	-
C-, B- = DIF against focal group; C+, B+ = DIF against reference group		
ETS (Educational Testing Service) use Delta δ units. 1 logit = 2.35 Delta δ units. 1 Delta δ unit = 0.426 logits.		
Zwick, R., Thayer, D.T., Lewis, C. (1999) An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. . Journal of Educational Measurement, 36, 1, 1-28 More explanation at www.ets.org/Media/Research/pdf/RR-12-08.pdf pp. 3,4		

9.5.2 Teachers' Leniency Measures on One Interval Scale

The new scale includes 20 items and 85 Chinese and 21 English teachers. The summary statistics are in *Table 78*. The average person's measure (MEAN) is -.11 logit. The person's measure is in a wide range of -3.63 logit to 2.09 logit with a standard deviation (P.SD) of 1.07 logit, indicating that most of the population spread in a good range of 2.14 logit on the scale. The reliability of the teachers' attitude estimates is .73, and the Separation is 1.62, which is relatively low. It might be due to a relatively large sample's limited number of items.

As shown in *Table 78,* the average (Mean) measure of items is located at 0 logit as default. The difficulty of items is in a wide range from -5.02 logit to 3.40 logit. Most item difficulties are located in a standard deviation of 2.75 logit. It indicates that items spread in a wide range of difficulties to measure the trait. Item reliability and separation indicate the ability of the measurement to define a distinct hierarchy of items along the measured trait. The higher the number, the more confident the study can apply the measurement across other samples and get the same order of item estimates. The item reliability is .99, and the separation is 9.12. They both are high. It suggests that the study can rely on this order of item estimates for other suitable samples.

Table 78 Summary statistics of persons and items

	Mean	P.SD	Min	Max	Separation	Reliability

Persons	-.11	1.07	-3.63	2.09	1.62	.73
Items	.00	2.75	-5.02	3.40	9.12	.99

The study had 106 teachers' leniency measures collected from China and England on the new scale. There is a significant difference in leniency measures between Chinese and English teachers (F (104) =2.55, p=.012, d=.62). The Chinese teachers' leniency measures (M=.02, SD=1.06) are significantly higher than the English teachers' measures (M=-.63 SD=.98). It indicates that Chinese teachers were more acceptable or more lenience towards these animated ADHD behaviours than English teachers.

9.6 Measuring Children's ADHD Symptoms across China and England

The study combined the Chinese and English samples of children's ratings on one scale and investigated DIF analysis to rule out items with potential bias. After calibrating the instrument, the study measured children's ADHD symptoms across China and England by anchoring teachers' leniency on one scale.

9.6.1 Calibrating the Instrument for Measuring Both Countries on One Scale

To investigate the invariance of the scale after combining the two samples, the study introduced two facets, gender and country, to the Many-Facet Rasch model to investigate DIF interaction with children's gender and country. However, these two facets do not involve in the estimates of the other three facets (teacher, children and behaviours) because their values were anchored as 0. Meanwhile, the teacher's leniency facet is anchored with the values that measured from the Teacher's Rating Scale of Cartoons. The analysis was conducted with the software Facets.

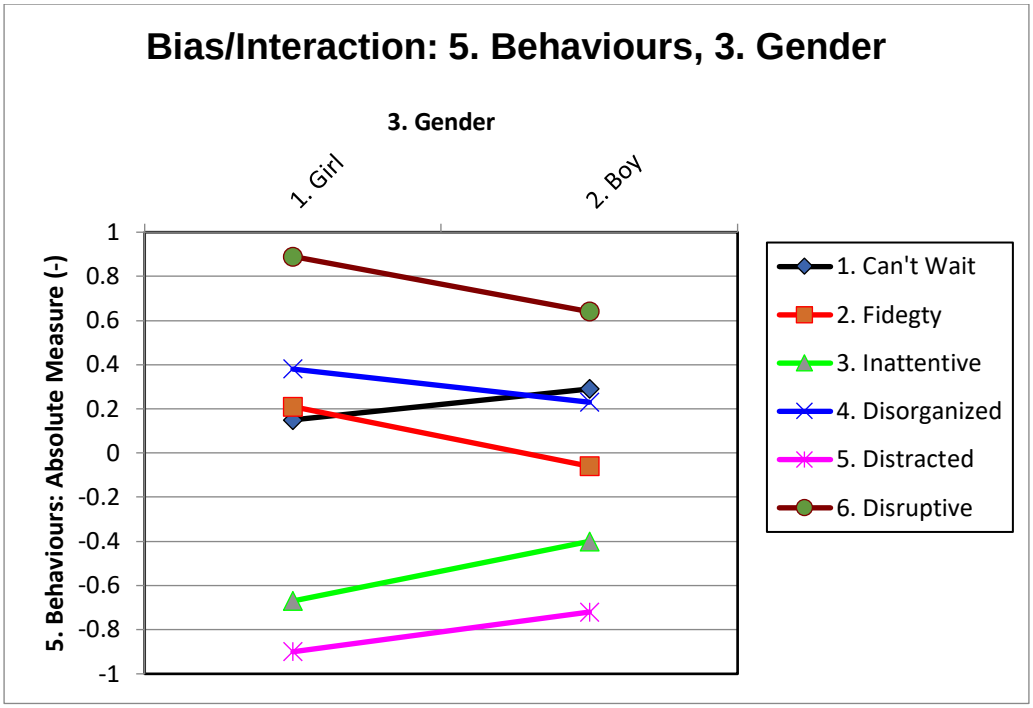
In *Table 79,* Target Contrast is the item difficulty difference between genders. There are three items (Fidgety, Disruptive and Disorganized) with positive Target Contrast, indicating that these items are more difficult for girls than boys. In other words, boys are rated higher on these behaviours than girls. Meanwhile, there are three items (Can't wait, Distracted, Inattentive)

with negative Target Contrast, which means that these items are more difficult for boys than for girls. In other words, girls are rated higher than boys on these behaviours. However, all Target Contrast values are low and negligible. Therefore, it indicates that the rating scale is invariant across gender. The DIF measures of six items across girls and boys are shown in Figure 61.

Table 79 DIF statistic for gender

DIF Measure (logit)		DIF Contrast (logit)	Prob	Items
Girls	Boys			
.15	.29	-.15	.20	Can't wait
.21	-.06	.26	.02	Fidgety
-.67	-.40	-.27	.02	Inattentive
.38	.23	.15	.23	Disorganised
-.90	-.72	-.18	.10	Distracted
.89	.64	.25	.04	Disruptive

Figure 61 DIF for gender



In *Table 80*, DIF Contrast is the item difficulty difference between China and England. There are four items (Inattentive, Disruptive, Can't wait and Distracted) with negative DIF Contrast, which indicates that these items are more difficult for children from England than children from China. In other words, Chinese children were rated higher than English children on these behaviours. There are two items (Fidgety and Disorganised) with positive DIF Contrast, which indicates that these items are more difficult for children from China than for children from England. In other words, English children were rated higher than Chinese children on these behaviours. *Figure 62* shows the differences of the two countries. However, there are only two items' (Inattentive and Disorganised) DIF values are significant ($p<.05$) and both differences are moderate. They are not big enough to raise concern. Therefore, it indicates that the rating scale is invariant across China and England.

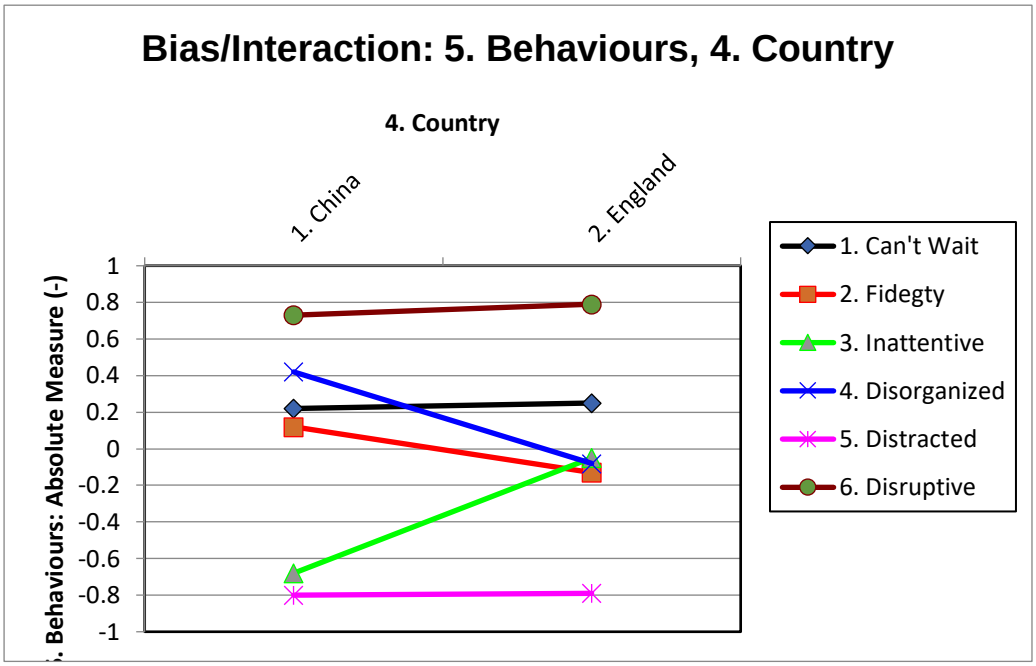
Table 80 DIF statistics for country

DIF Measure (logit)		DIF Contrast (logit)	Prob	Items
CN	ENG			

.22	.25	-.03	.82	Can't wait
.12	-.13	.25	.06	Fidgety
-.68	-.05	-.63	.0000	Inattentive
.42	-.08	.50	.0002	Disorganised
-.80	-.79	-.01	.91	Distracted
.73	.79	-.06	.67	Disruptive

Note. CN=China, ENG=England

Figure 62 DIF for country



9.6.2 Children's ADHD Symptoms Measures on One Interval Scale

In total, 760 children from both countries were assessed for six ADHD symptoms using a single equal-interval rating scale. Among them, there were 550 Chinese children, consisting of 256 girls and 294 boys. The descriptive data for these measures are presented in *Table 81*.

Additionally, there were 210 English children, comprising 99 girls and 111 boys, with their descriptive data for the measures displayed in *Table 82.*

Table 81Chinese children ADHD symptoms measures

China	N	Min	Max	Mean	SD
Total	550	-6.06	5.87	-.93	2.40
Girls	256	-6.06	5.87	-1.56	2.27
Boys	294	-6.06	5.87	-.38	2.37

Table 82 English children ADHD symptoms measures

England	N	Min	Max	Mean	SD
Total	210	-6.06	5.87	-1.42	2.87
Girls	99	-6.06	5.87	-2.22	2.70
Boys	111	-6.06	5.87	-.71	2.85

9.7 Adjusting Children's ADHD Symptoms Measures with Teachers' Leniency

In order to adjust children's ADHD behaviours measures by considering teachers' leniency, the study adopted Many-Facets Rasch Model to analyse three facets for estimating children's ADHD behaviours: item difficulty, children's behaviour and raters' leniency. In the three facets model, raters' leniency is normally estimated from their ratings of children. That requires teachers from both countries to cross-rate children from both countries, which is not able to conduct due to many reasons, as discussed in the previous Chapters. However, in the three facets model, the raters' leniency measures can be introduced as fixed values (anchored facet) with the values that were measured from the Teacher's Rating Scale of Cartoons. Since teachers from both countries rated animated ADHD behaviours, in this way, the study measured their leniency towards ADHD behaviours using the Teacher's Rating Scale of Cartoons. Therefore,

by anchoring teachers' leniency in the three facets analysis, the study managed to adjust children's measurements by considering teachers' leniency.

9.7.1 Anchoring Teachers' Leniency Measure in Many-Facet Rasch Model Analysis

However, before getting the children's ADHD symptoms measurements for both countries, the study investigated the reliability the Teacher's Rating Scale of Children's Behaviours with the combined children's samples and anchored teacher's leniency from both countries. The study explored Wright Map, Rater fit, behaviours and children's measurements for the combined samples and anchored facet (teacher leniency) with Facets software.

Wright Map. The Wright map of the three facets (children's behaviour, item difficulty and rater leniency) is shown in *Figure 63*. The first column is the logit scale labelled as 'Measr'. Three measures of teachers' leniency, children's ADHD symptoms and behaviour difficulty are positioned vertically on the vertical ruler with logit as the measurement units.

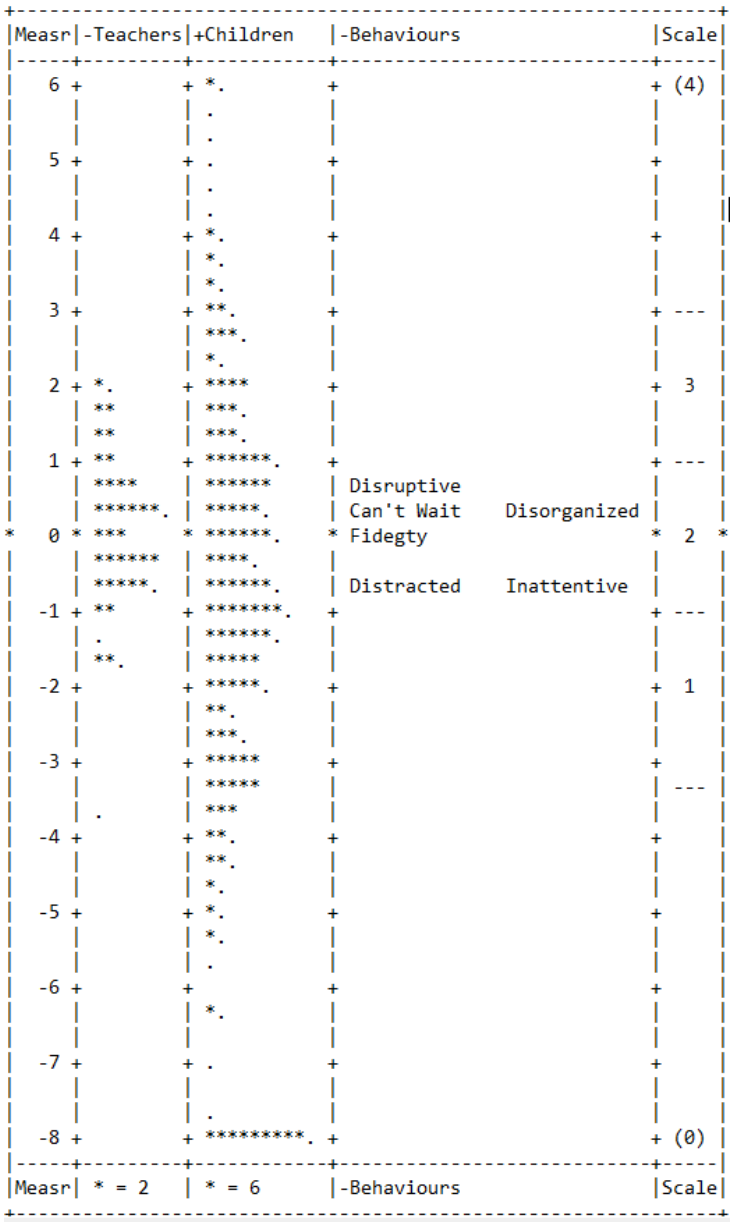
The second column, labelled '-Teachers', shows the teachers' leniency measures towards children's ADHD symptoms. More lenient teachers appear higher on the scale, and less lenient teachers appear lower on the scale. Since more lenient teachers tend to rate children's ADHD symptoms low, the teacher facet has a negative orientation with the children facet, indicated by the '- 'sign. As we can see from the scale, teachers' leniency varies by about 4 logit. About one-third (28.68%) of the logit spread observed for Children's ADHD symptoms measures (14 logit). The variability of teachers' leniency is substantial and would considerably impact the rating of children's symptoms.

The third column is labelled as "+Children" and shows the children's ADHD symptoms measures. Children ADHD symptoms measures are ordered with children who manifest higher

frequency symptoms at the top of the column and children who manifest lower frequency symptoms at the bottom of the column. Since the higher score the child has represents the high frequency of ADHD symptoms the child manifest, the facet is positively oriented on the scale with a" +" sign, as we can see that children are in a range of 6 logit to -8 logit.

The fourth column is labelled "-Behaviours" and presents the measures of six ADHD behaviours. This facet is negatively oriented with a "- "sign, which means that the behaviour appearing higher in the column is more difficult than the behaviour appearing lower in the column. As we can see, the "Disruptive" behaviour is at the top of the column, and it is the most difficult behaviour. The "Distracted" behaviour is at the bottom of the column, and it is the easiest behaviour. It also means that children scored higher on "Distracted" behaviour than "Disruptive" behaviour. The six behaviours spread in a relatively narrow range of 2 logit compared to other facets. It indicates that the difficulty across these six behaviours is not big. The last column is labelled "Scale" and maps the five-category scale (0, 1, 2, 3, 4) to the logit scale. The horizontal dashed lines in the scale column indicate the category threshold measures.

Figure 63 Wright map of three facet rating scale analysis.



Note. Each star in the second column represents two teachers, and a dot represents one teacher. Each star in the third column represents six children, and a dot represents one or two children.

Rater Fit. Some of the teacher's leniency measures, and fit statistics is shown in *Table 83*. As we can see that the column "Measure" are the measures of teacher's leniency. There are "A" characters in front of all the measures which indicate that these values are anchored or set. These measures are not estimated from teachers' ratings of children's ADHD behaviours. The teacher's leniency measures are values measured from the Teacher's Rating Scale of Cartoons.

The Infit and outfit columns are rater fit statistics. They indicate to what extent a given rater is associated with unexpected ratings. As Linacre (2008) suggested, the scale is productive for measurement if the infit and outfit mean-square values range from 0.4 to 1.6. Meanwhile, infit is commonly considered more critical than outfit in judging model fit, and misfit is generally more problematic than overfitting (Myford & Wolfe, 2003). As fit statistics orders in the table, we can see that the teachers on the top of the table are the most misfit raters. Misfitting raters reflect inconsistent rating behaviour. However, the rater misfit may also reflect possible idiosyncratic examinee performance in their ratings. While overfit raters indicate a central tendency or a halo effect in their rating behaviour (Myford & Wolfe, 2004; Engelhard, 2002).

Observed and fair averages give us a different view to interpret the leniency differences of raters in the raw score metric. The observed average is a teacher's mean rating across all children and behaviours that the teacher rated. The observed average does not tell us whether a high rating is due to the severity of the teacher or the high frequency of ADHD symptoms. However, fair average disentangles these two factors from the raw score metric. A fair average for a teacher k adjusts the observed average for the difference in the level of ADHD symptoms of the children teacher k rated from the children's ADHD symptoms mean across all teachers. A fair average enables a fair comparison between raters in the raw-score metric.

Table 83 Examples of rater leniency and fit statistics

Teacher	Leniency Measure	SE	Infit MNSQ	Outfit MNSQ	Fair Average	Observe Average	Number of Ratings
t079	A 2.09	.18	.72	.74	.73	2.00	60
t075	A 1.74	.19	.88	.82	.79	2.18	60
t059	A 1.39	.20	.96	.97	.93	1.10	60

t092	A .62	.20	1.02	1.01	1.27	1.78	60
t012	A -1.70	.18	.67	.72	2.54	2.05	60

Note. SE=Standard error. Infit MNSQ and outfit MNSQ are mean-square statistics.

Rater separation statistics are shown in *Table 84*. It indicates that the teachers are heterogeneous. The Fixed chi-squared statistics (Q=1553.6, df=75, p<0.01) reject the null hypothesis. It means that teachers' leniency measures in the population are not all the same. There are significant differences among this group of teachers. Separation gives the spread of the teacher leniency measures relative to the precision of those measures. For the sample data, separation is 4.36, indicating that teacher leniency differences were more than 4 times greater than the measurement error. For sample data, the strata are 6.15, indicating that the teachers can be separated into at least 6 statistically distinct groups by their leniency differences. Ideally, if all the teachers exercised a similar level of leniency, the separation should be close to value 0, and the strata should be close to value 1. As we can see in this sample, the teacher's leniency is widely spread on the scale, which means that the teachers are heterogeneous when they rate ADHD behaviours. It also indicates that it is essential to adjust children's ADHD behaviours according to teachers' leniency differences.

Table 84 Rater separation statistics

Separation	Strata	Reliability	Chi-squared
4.84	6.79	.96	1852.6 d.f=75 p<.001

Behaviours Measure Results. As we can see in *Table 85*, the difficulty measures of the six behaviours vary only in a small range (1.53 logit). Three behaviours' (Fidgety, Can't wait and Disorganised) difficulties are very similar. The most challenging behaviour is 'Disruptive', which indicates that most of the children were rated low on the behaviour by their teachers.

The easiest behaviour is 'Distracted', indicating that most of the children were rated high on this behaviour by their teachers.

Table 85 Behaviours measure results

Behaviours	Measure	Infit MNSQ	Outfit MNSQ	Correlation
Disruptive	.74	1.03	1.00	.82
Disorganised	.29	1.26	1.22	.81
Can't wait	.23	1.42	1.40	.78
Fidgety	.06	.79	.75	.87
Inattentive	-.52	.81	.79	.87
Distracted	-.80	.75	.76	.88

Children Measurement Results. Examples of children's estimates with anchored teachers' leniency from Facets software are shown in *Table 86.* In total, 760 children from both countries were assessed for six ADHD symptoms using a single equal-interval scale. Among them, there were 550 Chinese children, consisting of 256 girls and 294 boys. The descriptive data for these measures are presented in *Table 87*. Additionally, there were 210 English children, comprising 99 girls and 111 boys, with their descriptive data for the measures displayed in *Table 88*.

Table 86 Examples of children's measurements

Children	Behavioural Measure	SE	Infit MNSQ	Outfit MNSQ	Fair Average	Observe Average	Number of Ratings
cc366	4.52	1.07	.72	.50	3.83	3.83	6
cc434	2.88	.65	1.45	1.36	3.41	3.33	6
cc143	2.42	.57	2.02	2.00	3.22	2.83	6

ec094	.03	.55	.88	.86	1.95	2.33	6
ec101	-6.46	1.09	1.13	1.31	.03	.17	6

Note. SE=Standard error. Infit MNSQ and outfit MNSQ are mean-square statistics.

Table 87 Chinese children ADHD symptoms measures after anchoring teacher's leniency

China	N	Min	Max	Mean	SD
Total	550	-6.73	6.92	-.64	2.59
Girls	256	-6.73	6.92	-1.32	2.54
Boys	294	-6.73	5.83	-.04	2.49

Table 88 English children ADHD symptoms measures after anchoring teacher's leniency

England	N	Min	Max	Mean	SD
Total	210	-9.72	5.61	-2.06	3.15
Girls	99	-9.72	4.02	-2.91	2.96
Boys	111	-9.72	5.61	-.1.31	3.13

9.8 Investigating Changes in Children's Symptoms Measures after Anchoring Teachers' Leniency

The study compared children's measures before anchoring with the children's measures after anchoring teachers' leniency to investigate the differences. The study was interested in children's estimates changes rather than changes in individual item estimates. Therefore, the study only compared children's measures on the whole scale but did not look at individual items. The study conducted a paired sample T-test to compare children's measures before and after anchoring to investigate whether anchoring teachers' leniency has made significant changes in children's measures.

The paired sample T-test indicates that there is a significant difference in Chinese children's measures between before and after anchor (F (549) =-7.35, p<.001, d=-.31). Chinese children's measures before anchor (M=-.92, SD=2.40) have significantly increased after anchor (M=-.63 SD=2.60). This significant increase happened not only to Chinese girls' measures (F (255) =-4.19, p<.001, d=-.26) but also happened to Chinese boys' measures (F (293) =-6.12, p<.001, d=-.36). Chinese girls' measures before anchor (M=-1.56, SD=2.27) have significantly increased after anchor (M=-1.32 SD=2.54). Meanwhile, Chinese boys' measures before anchor (M=-.38, SD=2.37) have also significantly increased after anchor (M=-.04 SD=2.49). The results indicate that if we consider Chinese teachers' leniency in their ratings of Chinese children's ADHD behaviours, the children's measures significantly increased from without considering teachers' leniency.

However, the English sample tells a different story. The paired sample T-test indicates that there is a significant difference in English children's measures between before and after anchor (F (209) =9.69, p<.001, d=.67). English children's measures before anchor (M=-1.42, SD=2.87) have significantly decreased after anchor (M=-2.06 SD=3.15). This significant decrease happened not only to English girls' measures (F (99) =7.00, p<.001, d=.70) but also happened to English boys' measures (F (110) =6.69, p<.001, d=.64). English girls' measures before anchor (M=-2.21, SD=2.70) have significantly decreased after anchor (M=-2.91 SD=2.96). Meanwhile, English boys' measures before anchor (M=-.71, SD=2.85) have significantly decreased after anchor (M=-1.31 SD=3.13). The results indicate that if we consider English teachers' leniency in their ratings of English children's ADHD behaviours, the children's measures significantly decreased from without considering teachers' leniency.

Furthermore, the study found that the difference in ADHD behavioural measures between Chinese and English children has increased after anchoring teachers' leniency. Before anchoring teachers' leniency, Chinese children's measures (M=-.93 SD=2.40) are significantly higher (F (326) =2.22, p=.027, d=.20) than English children (M=-1.42, SD=2.87) by .49 logit. After anchoring teachers' leniency, Chinese children's measures (M=-.64, SD=2.59) are significantly higher (F (323) =5.86, p<.001, d=.52) than English children (M=-2.06, SD=3.15) by 1.41 logit.

Exploring Gender and Country Effects on Children's Measures After Anchoring

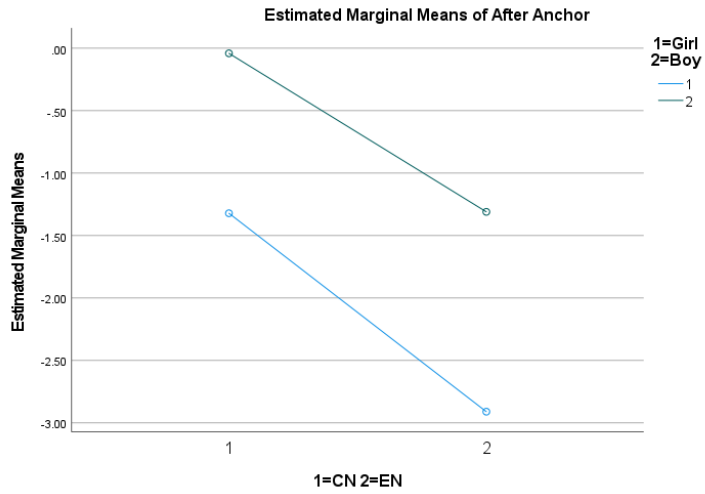
To understand children's measures after anchoring teachers' leniency, the study wanted to investigate whether gender and country these two factors have effects and interaction effects on children's ratings. To achieve that, we have conducted a Two-Way ANOVA analysis, and the statistics are shown in *Table 89*. First, there was a significant effect of the country factor for children's ADHD behavioural rating F (1,756) =43.37, p<0.001. The study found that Chinese children (M=-.64, SD=2.59) are rated significantly higher than English children (M=-2.06, SD=3.15) on their ADHD behaviours. Second, there was a significant effect of gender factor on children's ADHD behavioural rating F (1,756) =44.06, p<0.001. Boys (M=-.39, SD=2.73) were rated significantly higher than girls (M=-1.76, SD=2.75). However, there is no significant interaction effect of country and gender on children's ratings F (1,756) =.54, p=.46. As we can see in *Figure 64*, the green line (boys) is higher than the blue line (girls), and the two lines are almost parallel. It indicates no interaction effect of country and gender factors on children's behavioural ratings.

Table 89 Measures for gender and country

Country	Gender	Mean	SD	N
CN	Girls	-1.32	2.54	256
	Boys	-.04	2.49	294

	Total	-.64	2.59	550
EN	Girls	-2.91	2.96	99
	Boys	-1.31	3.13	111
	Total	-2.06	3.15	210
Total	Girls	-1.76	2.75	355
	Boys	-.39	2.73	405
	Total	-1.03	2.83	760

Figure 64 Gender difference across two countries



9.9 Conclusion

In conclusion, this chapter has presented the results of a comprehensive research study aimed at developing measurement instruments and investigating rater bias in assessing Attention-Deficit/Hyperactivity Disorder (ADHD) behaviours among teachers from different cultural backgrounds in China and England. The primary objective of this study was to address the challenges posed by subjective judgments of children's ADHD behaviours, influenced by cultural and social environments. The results of our study indicated significant differences in the tolerance levels of teachers from different cultural backgrounds towards children's ADHD behaviours. This finding suggests that cultural factors can influence the way teachers perceive

and rate these behaviours, which may impact the overall assessment of ADHD in children. This highlights the importance of developing culturally sensitive measurement instruments and methods to ensure fairness and accuracy in ADHD assessments across different cultural contexts. The next chapter will delve into a detailed interpretation of the results, providing deeper insights into the interplay between cultural influences, rater bias, and the assessment of ADHD behaviours among teachers in China and England. Through an in-depth analysis, we aim to gain a comprehensive understanding of the implications of the study findings and shed further light on the complexities of ADHD assessments in diverse cultural settings.

Chapter 10 Discussion

10.1 Introduction

The preceding chapters have presented a comprehensive examination of the research findings and methodologies concerning the assessment of children's ADHD behaviours in cross-cultural contexts. In this chapter, I delve into a detailed discussion of the key findings, implications, and contributions of the study. The primary objective of this research was to develop a robust and fair approach to measuring and adjusting for rater leniency in the assessment of ADHD behaviours across diverse cultural groups. The innovative use of cartoon animations and the development of the Teacher's Rating Scale of Cartoons (TRSC) have proven to be invaluable tools in enhancing the reliability and validity of ADHD behaviour assessment. By utilising the Many-Facet Rasch Model (MFRM) with anchored leniency measures from TRSC, we were able to overcome the challenges associated with cross-rating in a cross-country study, ultimately allowing for fair and accurate comparisons between children from China and England. This chapter aims to provide a comprehensive synthesis of the research journey, highlight the significance of our findings, and offer recommendations for future research and practice in cross-cultural assessments of children's ADHD behaviours.

10.2 Research Question 1

How to measure the leniency of Chinese and English teachers in rating children's ADHD behaviours?

10.2.1 An Innovative Approach

Rater leniency, the tendency of raters to consistently assign higher or lower ratings compared to their peers regardless of the actual level of observed behaviours, poses a challenge in measurement reliability and fairness. It is imperative to quantify and address rater leniency to enhance standardisation and ensure fairness in assessments. In cross-cultural research, understanding and accounting for rater leniency become even more critical due to cultural

variations in behaviour perception and interpretation, leading to discrepancies in ratings across different cultural groups. By measuring and adjusting for rater leniency, the comparability of scores can be improved, ensuring that observed differences across cultures truly reflect variations in the measured construct. Ultimately, measuring rater leniency facilitates fair comparisons between different cultural groups, particularly in cross-cultural contexts.

This study represents a pioneering effort in developing a rating scale that employs a series of engaging cartoon animations as rating objects to assess teachers' leniency towards children's ADHD behaviours in two distinct countries. When conducting cross-cultural studies on teachers' leniency, the first challenge pertains to selecting a method for developing rating objects that can be effectively utilized across different cultural contexts, enabling teachers from diverse cultural backgrounds to observe and rate behaviours consistently. Unlike previous studies employing written vignettes or video recordings, this study creatively utilized cartoon animations to portray children's ADHD behaviours within a classroom setting, allowing teachers from China and England to observe and rate these behaviours. To accompany the animations, a questionnaire utilizing the Semantic Differential Scale was developed, enabling teachers to indicate their level of acceptance of observed animated behaviours. This innovative approach facilitated the measurement of teachers' leniency towards children's ADHD behaviours across the two countries.

Developing accurate cartoon animations that depict children's ADHD behaviours involves two crucial stages: writing cartoon scripts and converting them into animations. To ensure high-quality results throughout these stages, this study employed a creative multi-method approach to creating a comprehensive ADHD behavioural bank consisting of 18 symptoms. This bank served as valuable material for scriptwriting, enabling the accurate portrayal of ADHD

behaviours in the cartoons. To maintain consistency and uphold quality during the conversion of scripts to animations, the study collaborated with a team of professional animators and followed standard industrial procedures to produce animations.

Cartoon scripts are crucial for creating cartoon animations because they serve as the foundation and blueprint for the entire animation process. They provide structure, direction, and coherence to the storytelling process. They serve as a guide for animators, aid in character development, and facilitate effective communication and collaboration within the animation team. A well-written script lays the foundation for a successful and engaging cartoon animation.

To ensure the validity of the scripts, the study designed a multi-method approach including classroom observation, teacher interview and rating scales review to create a children's ADHD behavioural bank for all the symptoms. Classroom observation was conducted in both English and Chinese schools to understand the context in which ADHD behaviours occurred. This naturalistic observation provided first-hand information about children's ADHD behaviours, settings, and routines in the classrooms. The observation data, recorded in observation diaries, captured the behaviours and related contexts. Additionally, interviews were conducted with Year 2 teachers from both countries to gain their perspectives and opinions on ADHD symptoms and behaviours in their classrooms. The interviews helped validate the observed behaviours and provided insights into the teachers' interpretations. The behaviours teachers mentioned in the interview were extracted into a table which kept all the children's behaviours for each symptom in the interview. The scripts were also informed by the review of rating scales that are commonly used to collect teachers' ratings of children's ADHD behaviours. The review focused on both broad-band and narrow-band rating scales, which included a wide range of behavioural and emotional problems related to ADHD symptoms. By understanding

the behaviours teachers were asked to observe in these rating scales, the scripts could incorporate relevant and representative ADHD behaviours. All the behaviours that the study gathered from the three methods served as a behavioural bank for generating scripts to portray a wide range of children's behaviours of each ADHD symptom in various contexts.

By incorporating insights from classroom observation, teacher interviews, and rating scale reviews, the scripts for the cartoons aimed to accurately portray children's ADHD behaviours in a valid and reliable manner. The multi-method approach helped address potential biases and ensured that the scripts were grounded in real-world observations and teachers' perspectives.

In order to maintain the consistency between the scripts and animations to develop a series of standard animations, this study worked with a professional team and followed standard industrial procedures to produce animations. The procedure involves transforming scripts into visual representations through storyboarding, background design, character design, and animation production.

Storyboard Creation: Once the scripts reach their final draft, a storyboard is created. The storyboard is a visual interpretation of the script, consisting of small thumbnail drawings that depict every scene and indicate action and camera moves. It helps set up the scenes properly and adds dramatic elements to enhance storytelling.

Background Design: Background sketches are developed based on the descriptions of the scenarios in the scripts. Since the stories take place in a classroom, the background design includes sketches of the classroom from different camera angles to provide a suitable setting.

Character Design: The characters in the cartoons include a teacher and several children. The character design involves determining the appearance of the characters, such as their skin colour, facial features, and clothing. The characters should resemble 6 to 7-year-olds without specific cultural references or gender indications.

Storyboard Revision: The storyboard goes through a revision process with input from the team to ensure clarity, continuity, and creative storytelling. Once the revisions are finalized, the production moves to the next step.

Animation Production: The final storyboard serves as a visual guide for the animators who bring the illustrations to life using specialized software. They add movements to the characters and other elements in the cartoons. Background music is also incorporated to enhance the viewing experience. The process often involves multiple revisions to align the animations with the scripts and ensure a cohesive final product.

The two stages (writing scripts and producing animations) undergo a review process involving two PhD supervisors, a Year 2 English teacher and a Year 2 Chinese teacher, to ensure clarity and understanding. PhD Supervisors, as well as Chinese and English teachers, bring their expertise and experience in education and classroom settings. Their involvement ensures that the scripts and animations accurately depict the behaviours and contexts relevant to ADHD in the classroom. Their insights help maintain the educational validity and authenticity of the content, ensuring that it aligns with real-world teaching experiences. Teachers are the primary audience for the cartoons, as they are intended to measure their attitudes toward ADHD behaviours. Involving them ensures that the content is relevant, relatable, and meaningful to their professional context. Their involvement allows for iterative feedback and revisions based

on their perspectives, resulting in a final product that effectively meets their needs and requirements.

10.2.2 Traditional Approaches

Prior research has examined teachers' biases when rating children's ADHD behaviours, often employing written vignettes as a methodology. For instance, Lane et al. (2008) utilized written vignettes to investigate how teacher characteristics and student characteristics influenced teachers' perception and recognition of disruptive behaviours. The vignettes described students displaying disruptive behaviours, and teachers were asked to rate the severity of the behaviours and suggest intervention strategies. By employing written vignettes, the researchers ensured consistent and standardised scenarios for eliciting teacher responses, enabling a systematic examination of the factors influencing their perceptions and evaluations. Similarly, Owens et al. (2017) employed written vignettes to examine the role of race and gender in teachers' recognition of ADHD. Teachers were presented with vignettes describing children exhibiting ADHD symptoms and were asked to evaluate and make referral decisions based on the information provided. The use of written vignettes allowed for controlled comparisons of teacher responses, shedding light on factors such as race, gender, stereotypes, and diagnostic labels and their potential impact on teacher evaluations of student behaviours.

However, text vignettes have limitations for this study. They may not capture the full range of behaviours and contexts associated with ADHD symptoms, as conveying complex behaviour patterns through written descriptions can be challenging. The visual and observational richness of presenting behaviours directly or through video recordings is lost in text vignettes. Additionally, the standardised nature and limited length of text vignettes may restrict the inclusion of detailed information about the child's behaviours and contexts, potentially leading to gaps in understanding. The translation of text vignettes for cross-cultural studies raises

concerns about their equivalence and standardisation across different languages. Moreover, people's perceptions and judgments of ADHD behaviours are often based on real-life observations rather than reading about them in vignettes, which may create a disconnect between responses to vignettes and real-world situations. Therefore, alternative methods that incorporate real-life observations or more comprehensive assessments, such as video recordings or cartoon animations, may be more suitable for studying cross-cultural and cross-linguistic differences in teachers' perceptions or attitudes towards ADHD symptoms.

Video tapes have been widely used in studies to simulate behaviours and symptoms associated with psychiatric disorders, such as ADHD. These tapes provide a standardized display of the disorders, allowing for controlled experimental conditions. For example, Jackson and King (2004) utilized video tapes to explore the impact of gender differences and oppositional behaviour on teachers' ratings of ADHD symptoms. They created six videotapes depicting the behaviours of children with ADHD, Oppositional Defiant Disorder (ODD), and typically developing children. The tapes were developed based on established guidelines and previous studies' transcripts. Careful attention was given to maintaining consistency in the behaviour of teachers and other children in the classroom. Child actors were recruited, and the same actor portrayed both ADHD and ODD behaviours. Filming took place in a real 4th-grade classroom, and an experienced consultant directed the process.

Similarly, Paulson et al. (2005) used video tapes to examine the influence of ADHD, depressive, and anxiety symptoms on social rejection and mood states. They created five different segments, each featuring an actress portraying behaviours characteristic of a specific disorder. The scripts were designed to elicit the target symptom presentations based on diagnostic criteria. The video tapes, approximately 20 minutes long, showcased the actress interacting with the

camera in a neutral background, simulating conversations and responses to questions. Multiple rehearsals and takes were conducted to ensure accurate portrayal and independent raters assessed the believability and diagnostic accuracy of the video segments.

The use of video tapes in these studies offers several advantages. Firstly, they provide standardized and controlled stimuli to examine the impact of specific behaviours and symptoms. Secondly, video tapes allow for the observation of behaviours in realistic settings, providing a more authentic representation of social interactions. Consistency in behaviour across tapes is maintained through careful management, reducing confounding factors. The involvement of trained actors and adherence to diagnostic criteria enhances the accuracy and credibility of the portrayed behaviours. Additionally, the review and rating of the video segments by independent raters ensure the reliability of the collected data.

However, in the context of cross-cultural studies, video tapes may introduce potential cultural bias due to factors such as language and physical appearance. To address this, cartoon animations offer advantages in controlling potential cultural bias. They provide a more efficient medium for illustrating ADHD behaviours, demonstrate a bigger picture than written descriptions, and eliminate potential biases introduced by language translation. Cartoons also limit cultural bias in assessment instruments, as animated characters can be designed with fewer cultural and gender references. Moreover, cartoons allow for the creation of mutual classroom settings and activities, accommodating cultural differences and avoiding confusion caused by specific country contexts. Their ease of production, flexibility, and ability to engage informants further contribute to their suitability for this type of study.

10.2.3 Differences in Teachers Ratings and Factors Influence the Ratings

With the cartoon animation media, the study developed a measurement instrument called the Teachers' Rating Scale of Cartoons, which consisted of six episodes of cartoon animations representing different symptoms of ADHD. The symptoms included Can't wait, Fidgety, Inattentive, Disorganised, Distracted, and Disruptive. Each episode depicted six behaviours, with five behaviours related to ADHD and one representing typical behaviour for each symptom. Overall, the scale contained 30 animated behaviours that teachers were required to observe and rate.

To examine the effectiveness of the rating scale, the study involved a sample of 109 Chinese Year 2 teachers and 23 English Year 2 teachers who watched the cartoon animations and provided their ratings for the depicted behaviours. Rasch analysis was employed to analyse the data and evaluate the functioning of the rating scale for the respective samples. The Teacher's Rating Scale of Cartoons (TRSC) has shown good reliability and function in measuring teachers' leniency toward ADHD behaviours in both samples. The scale demonstrated acceptable category functioning, indicating that the response categories effectively measured the sample. Summary statistics of persons and items indicated that the scale effectively discriminated between individuals with different levels of ability, with high reliability and separation. The Wright Map revealed a reasonable alignment between persons' abilities and items' difficulties, although additional items could enhance the discrimination of persons' abilities.

The scale exhibited an overall good fit to the Rasch model, with most items falling within acceptable ranges for INFIT and OUTFIT mean-square values. There were no negative correlations between item responses and abilities, indicating that all items aligned with the

construct being measured. The scale demonstrated good unidimensionality, although there was a potential sub-dimension related to focusing behaviour. However, this sub-dimension did not form a distinct dimension in the measurement. In conclusion, the Teacher's Rating Scale of Cartoons (TRSC) is a reliable and functional instrument for assessing teachers' leniency toward ADHD behaviours in both cultural contexts. The scale effectively discriminates between individuals with different levels of ability and provides valuable insights into the perception and evaluation of ADHD behaviours. Further refinement of the scale, including the addition of items and exploration of potential sub-dimensions, could enhance its precision and applicability in measuring teachers' leniency in various settings.

Using the scale, the study found significant differences in how teachers from both countries rated 13 behaviours across the six symptoms. Chinese teachers rated six behaviours significantly higher than English teachers in the symptoms 'Can't wait,' 'Inattentive,' and 'Disruptive,' indicating a higher tolerance among Chinese teachers for these behaviours. On the other hand, English teachers rated five behaviours significantly higher than Chinese teachers in the symptoms 'Disorganised' and 'Distracted,' showing a greater tolerance among English teachers for these behaviours. In the symptom 'Fidgety,' Chinese teachers were more tolerant of the behaviour of 'moving body on the seat,' while English teachers were more tolerant of the behaviour of 'fiddling with other's hair.' These findings reveal that almost half of the animated behaviours showed significant rating differences between teachers from the two countries, suggesting that teachers from different cultures perceive children's ADHD behaviours differently. It is important to take these rater differences into account when comparing ADHD behaviours across cultures to avoid confounding results.

Numerous cross-cultural studies consistently highlight differences among raters in their evaluations. Du et al. (2003) discovered that teachers from Guangzhou rated significantly higher than teachers from Leeds and Hong Kong in two behaviours: leaving the seat and impatience when waiting for their turn. Conversely, teachers from Leeds rated significantly higher than Guangzhou teachers in the behaviour of taking someone else's belongings without permission. In another study, Alban-Metcalfe et al. (2002) found that Mainland Chinese teachers rated inattentive, hyperactive, and impulsive behaviours significantly higher than teachers from Hong Kong and the UK. Similarly, Mueller et al. (1995) examined disruptive behaviour ratings by teachers from five countries and observed that American and Japanese teachers rated the behaviours as less extreme compared to Chinese, Indonesian, and Thai teachers. Norvilitis and Fang (2005) also found that Chinese teachers tended to rate inattentive behaviours lower and hyperactive behaviours higher than their American colleagues. These studies, including my research, clearly demonstrate cultural discrepancies in the ratings of ADHD behaviours by raters from different cultural background.

Meanwhile, we also examined various factors related to teacher backgrounds that may influence their ratings, including gender, teaching position, teaching experience, and class size. Regarding gender, it is important to note that there were few male teachers working in primary schools in both countries, resulting in an insufficient number of male teachers in our sample. As a result, we were unable to determine whether there is a significant difference in ratings between female and male teachers in both countries. However, previous studies have found that male teachers tend to be more tolerant and hold more positive views about children's problematic behaviours and achievements compared to their female colleagues (Borg & Falzon,1990; Wheldall, 1992; Mullola et al., 2012). Collecting data from more male teachers would be valuable for further investigation in our study. In terms of teaching positions, we did

not find any significant differences in ratings among teachers with different positions in both countries. This suggests that teaching positions may not significantly influence teachers' ratings of children's ADHD behaviours.

Regarding teaching experience, the study observed negative correlations between teaching experience and ratings of certain behaviours in both countries. Chinese teachers' teaching experience showed weak negative associations with four ADHD behaviours that they rated, while English teachers' teaching experience demonstrated five moderate negative correlations with five ADHD behaviours they rated. These findings suggest that the more teaching experience teachers have or the older they are, the more likely they are to be harsher or less tolerant of some ADHD behaviours in their classrooms. Older teachers may be more influenced by traditional perceptions of children and have higher expectations regarding discipline compared to younger teachers. It may also reflect the fact that experienced teachers have been exposed to societal norms and expectations of children for a longer duration, making them more aware of the negative impact these behaviours can have on children's academic achievement. These findings align with the study by Alban-Metcalfe et al. (2002), where teachers rated children's ADHD behaviours significantly higher than student teachers when observing the same child. Mulholland et al., (2015) found that as a teacher becomes more experienced, he or she also becomes less tolerant of disruptive behaviours and less sympathetic to students who exhibit ADHD-type behaviours. The findings of these studies indicate that there are negative correlations between teaching experience and ratings of certain ADHD behaviours.

However, it is worth noting that Borg and Falzon (1990) found that more experienced teachers viewed problematic behaviours less seriously than teachers with less experience, suggesting

some variations in findings across studies. Schultz and Evans (2012) also found that women teachers and younger teachers tended to provide more severe ratings than men or older teachers for hyperactivity-impulsivity. Moreover, an increase in teacher age predicted greater leniency in hyperactivity-impulsivity ratings. This may be attributed to the older teachers' better classroom management skills and the establishment of well-structured classroom environments. Another possible explanation is that older teachers may make judgments based on comparisons with previous cohorts of students, making certain hyperactive behaviours appear less extreme. In summary, the various findings regarding teaching experience and ratings of ADHD behaviours can be influenced by factors such as traditional perceptions of children, expectations regarding discipline, exposure to societal norms, cultural context, and individual differences among teachers. The age and experience of teachers, as well as their classroom management skills and comparisons with previous cohorts of students, may play a role in shaping their evaluations and tolerance levels for ADHD behaviours.

Regarding class size, in the Chinese sample, we did not find any convincing correlation between teachers' ratings and class size. However, in the English sample, we observed a moderate positive correlation between English teachers' ratings and class size. This indicates that English teachers who taught relatively larger classes tend to be more tolerant of shouting out behaviour compared to teachers who taught smaller classes. One possible interpretation of these findings is that teachers with larger class sizes may encounter greater challenges in managing student behaviour and maintaining classroom order. With a higher number of students to attend to, teachers may prioritize addressing other disruptive behaviours or focus on overall classroom control, leading to a higher tolerance for shouting out behaviour. On the other hand, teachers with smaller class sizes may have more individualized attention and better control over classroom dynamics. With fewer students to manage, they may have more time

and resources to address and discourage shouting-out behaviour, resulting in a lower tolerance for such behaviours. While there is no specific study on how class size affects teachers' leniency towards children's ADHD behaviours in the classroom, other research has indicated that teachers' workload and stress levels can influence their leniency towards children's disruptive behaviours. For instance, Kokkinos et al. (2004) found that teachers with high workloads and stress levels tend to be less tolerant of children's disturbing behaviours.

In summary, the study investigated various factors that may influence teachers' ratings, including gender, teaching position, teaching experience, and class size. While some factors showed significant associations with teachers' ratings, further research is needed to delve deeper into these relationships and better understand the underlying mechanisms.

10.2.4 Teachers' Leniency Measures

The current study developed the Teachers' Rating Scale of Cartoons (TRSC) to measure teachers' leniency towards ADHD behaviours in China and England. Through the application of Rasch Model analysis, the study combined samples from both countries to calibrate the instrument and estimate teachers' leniency measures. DIF analysis was employed to identify potential sources of bias to calibrate the instrument, resulting in the exclusion of certain behaviours from the instrument to ensure equitable assessment across the two teacher groups.

The measures from the Teacher's Rating Scale of Cartoons indicated that there is a significant difference in leniency measures between Chinese and English teachers ($F(104) = 2.55$, $p = .012$, $d = .62$). The Chinese teachers' leniency measures ($M = .02$, $SD = 1.06$) are significantly higher than the English teachers' measures ($M = -.63$ $SD = .98$).This discrepancy suggests that Chinese teachers may exhibit a greater degree of tolerance and acceptance towards animated ADHD

behaviours. Several factors, including differences in the education system and cultural influences, could contribute to this divergence.

The study findings shed light on the influence of different teaching styles on teachers' tolerance towards ADHD behaviours. In China, classrooms typically follow a more teacher-centred approach, where the main focus is on delivering knowledge and supervising students' independent practice. Discipline or intervention for misbehaviours is given less priority unless they disrupt the overall learning environment. As a result, teachers may divert their attention away from ADHD behaviours unless they significantly impact the classroom dynamics. In contrast, in England, classrooms adopt a student-centred approach with an emphasis on activities and group work. Given that ADHD behaviours can not only affect the individual's learning but also disrupt other students' engagement, English teachers exhibited less tolerance towards such behaviours. These differences in teaching styles provide valuable insights into how classroom practices may shape teachers' responses to ADHD behaviours.

Furthermore, cultural factors play a significant role in shaping teachers' attitudes towards ADHD behaviours. In Chinese culture, academic success is often highly valued, and parents prioritize the achievement of good grades. This emphasis on academic excellence may lead Chinese teachers to be more accepting and accommodating of ADHD behaviours, as long as they do not hinder overall academic progress. On the other hand, in England, schools place importance on a broader set of values, including democracy, the rule of law, individual liberty, and mutual respect and tolerance for diverse beliefs. As a result, teachers in England not only focus on academic achievement but also place greater attention on encouraging positive behaviours that align with these values. This cultural context leads to a heightened awareness

and emphasis on children's behaviours in the classroom, including the management of ADHD behaviours.

Other studies have contradictory findings. Alban-Metcalf et al. (2002) found that Chinese teachers rated ADHD behaviours significantly higher than their colleagues from Hong Kong and Leeds, even observing the same child. It shows that Chinese teachers are less accepting of children's ADHD behaviours than their English colleagues. In a study comparing teachers' ratings of children's disruptive behaviour in five countries, Chinese teachers are harsher than their American and Japanese colleagues (Mueller et al., 1995). However, none of these studies demonstrated any information about the measurements they used for their cross-cultural comparison worked accurately and equally for different groups of participants. Therefore, the differences these studies found may also be due to measurement bias. Additionally, previous cross-cultural studies involving Chinese samples were conducted several decades ago, prior to significant societal changes and the incorporation of Western knowledge on child psychology and development into Chinese teacher training programs. These changes may have influenced Chinese teachers' attitudes towards children's behaviours over time.

It is important to note that this study focused on a limited set of ADHD behaviours and symptoms, which may not capture the entirety of the construct. The ADHD diagnostic criteria encompass a broader range of symptoms and behaviours across diverse contexts. Therefore, caution should be exercised in generalizing these findings to all Chinese or English teachers. Individual differences and variations within cultural contexts can also contribute to variations in teachers' attitudes. Further research and exploration of these factors would contribute to a more comprehensive understanding of why Chinese teachers, as observed in this specific study,

demonstrated greater leniency towards animated ADHD behaviours compared to English teachers.

10.3 Research Question 2

How to make appropriate adjustments to ratings of children's ADHD behaviours to improve fairness in cross-cultural comparison?

Before proceeding with appropriate adjustments to children's ADHD behaviours, the study developed a specialized instrument, the Teacher's Rating Scale of Children's Behaviours (TRSCB), to measure children's ADHD symptoms in both China and England. The TRSCB encompasses six ADHD symptoms, namely "Can't wait," "Fidgety," "Inattention," "Disorganised," "Distracted," and "Disruptive." These symptoms were derived from the Teacher's Rating Scale for Cartoons, which was used to assess teachers' leniency in both countries. With the TRSCB, participating teachers from China and England randomly selected 10 children from their respective classes and rated the behaviours of these selected children based on the six identified symptoms.

The study employed the Rasch Model analysis to examine the effectiveness of the scale in measuring children from diverse cultural backgrounds. The results demonstrated that the Teacher's Rating Scale of Children's Behaviours (TRSCB) proved to be a reliable instrument for assessing ADHD behaviours in both the Chinese and English samples. The category function analysis indicated that the scale effectively measured the samples, with INFIT and OUTFIT mean-square values falling within acceptable ranges. Additionally, the summary statistics for children and symptoms revealed that the scale successfully differentiated individuals with varying levels of ADHD severity, exhibiting high reliability and separation. Furthermore, the misfit statistics demonstrated that all symptoms aligned with the measured ADHD behaviours, and there were no negative correlations between symptom responses and

the disorder. The scale displayed good unidimensionality, with no meaningful subdimensions, and no issues of local dependency were observed. Based on these findings, the TRSCB was a robust instrument for assessing children's ADHD behaviour through teachers' ratings.

Given the scale's efficacy in measuring children's ADHD behaviours in both China and England, the study proceeded to combine the two samples to assess the behaviours of both groups on a unified scale. Employing Differential Item Functioning (DIF) analysis for country and gender variables, the study investigated whether the scale functioned equally for children of different genders and from different countries. Remarkably, the analysis revealed that the scale maintained consistent performance in measuring various groups of children, regardless of their gender or country of origin. These findings further confirmed the scale's cross-cultural applicability and highlighted its ability to provide fair and accurate assessments of ADHD behaviours in diverse populations.

After carefully assessing children's ADHD behaviours across China and England, the study employed the Many-Facet Rasch Model (MFRM) to make necessary adjustments to the children's measures considering teachers' leniency. This approach has been widely utilized in various studies to effectively adjust measures for rating scales. For instance, Farrokhi et al. (2012) used MFRM to adjust measures of students' English writing ability by considering three types of raters: teachers, self-assessors, and peer-assessors. The study successfully applied MFRM to adjust the measures of English writing abilities, accounting for rater severity, rater type, and task difficulty. In another study, Primi et al. (2019) utilized MFRM to adjust measures of creativity, taking into account rater leniency. Participants generated two metaphors that were rated by three raters on a 5-point creativity scale. Similarly, Wesolowski et al. (2016) applied MFRM to adjust measures for music performance using a scale of 22 items adapted from the Jazz Big Band Performance Rating Scale, which were evaluated by expert judges across

various performance levels. The MFRM analysis allowed for adjusting the performance measures with judges' leniency, performance level, and rating criteria difficulty. By employing MFRM, the analysis effectively adjusted ability measures, accounting for the variance of other facets such as raters, tasks, and time within the study.

The issue of rater leniency/severity variation is a common bias in rater-mediated rating scales. MFRM has proven effective in addressing this bias, especially in high-stakes scenarios, such as when raters assign scores to essays for college admissions tests (Eckes, 2015; Wind & Peterson, 2018). MFRM typically estimates rater leniency/severity through cross ratings, which necessitate special rating designs. The ideal design is a complete block design, also known as full cross-rating, where each examinee is rated by all examiners on all tasks. While complete data generated from full cross-rating leads to the most accurate estimates for all facets, it can be time-consuming and resource-intensive, particularly in situations with a large number of responses to be evaluated by multiple raters. To address this, most studies adopt a balanced incomplete block design, where raters are divided into groups, and each group evaluates only a subset of responses. This design ensures that each response is rated by a subset of raters, allowing for the estimation of measures for rater leniency/severity in MFRM analysis. Studies like Primi et al. (2019) and Wesolowski et al. (2016) have successfully employed this rating design, enabling MFRM to adjust measures for rater leniency/severity effectively.

However, cross-rating is not a practical option for conducting cross-country studies due to the logistical challenges of having Chinese and English teachers travel to each other's countries and spend extensive time observing and rating children's behaviours. As a result, the Many-Facet Rasch Model (MFRM) cannot directly estimate teachers' leniency from their ratings since each child was rated by only one teacher, leading to a lack of overlap in the ratings that would serve as references for MFRM to assess relative rater leniency. Nevertheless, measures can be

anchored individually or by group mean, allowing for equating and linking across test sessions. To address this limitation, the study devised the Teacher's Rating Scale of Cartoons (TRSC) to measure teachers' leniency separately. The leniency measures obtained from TRSC were then introduced as fixed values into the MFRM analysis of TRSCB (children's ADHD behaviour measures) to serve as teachers' leniency (fixed facet) estimates. Consequently, MFRM effectively adjusted children's ADHD behaviour measures according to their respective rater's leniency measures.

Anchoring facet values in MFRM analysis is a widely used approach to identify facet bias and make necessary adjustments to the measures. For instance, in Hung et al.'s (2012) study, all items were anchored at the same difficulty level to identify raters who exhibited the "halo" effect in their ratings. Similarly, Aryadoust (2016) employed anchored values of instructors' evaluations for students' presentations in MFRM analysis to investigate gender and academic subject bias in peer assessment. Engelhard (1996) introduced anchored values of expert panel evaluations for high school graduation writing ability to examine the accuracy of a group of raters.

In conclusion, the present study developed the Teacher's Rating Scale of Children's Behaviours (TRSCB) to measure children's ADHD behaviours in China and England. The scale demonstrated good reliability and effectively discriminated between individuals with different levels of ADHD severity. To address the common bias of rater leniency in rating scales, the study utilized the Many-Facet Rasch Model (MFRM). However, due to the impracticality of cross-rating for this cross-country study, the researchers innovatively used anchoring techniques, separately measuring teachers' leniency through TRSC and introducing these measures as fixed facets in the MFRM analysis of TRSCB. This approach provided a different

approach to assess and adjust children's ADHD behaviour measures, ensuring fair and accurate cross-cultural comparisons of ADHD behaviours.

10.4 Research Question 3

To what extent does the rater bias affect the ratings of children's ADHD behaviours in England and China?

The Teacher's Rating Scale of Cartoons revealed a significant difference in leniency measures between Chinese and English teachers ($F(104) = 2.55$, $p = 0.012$, $d = 0.62$). Chinese teachers' leniency measures ($M = 0.02$, $SD = 1.06$) were significantly higher than those of English teachers ($M = -0.63$, $SD = 0.98$), indicating a greater acceptance or leniency towards animated ADHD behaviours among Chinese teachers.

Before incorporating teacher's leniency measures to adjust children's ADHD behavioural measures, a comparison of children's measures between the two countries showed that Chinese children's ADHD behavioural measures were significantly higher than those of English children by 0.49 logit unit. After implementing teacher leniency measures, significant changes were observed in the children's measures for both countries, but in opposite directions. Chinese children's measures increased significantly by 0.29 logit after anchoring teachers' leniency, while English children's measures decreased significantly by 0.64 logits after the same adjustment. Consequently, the difference in children's measures between the two countries widened from 0.49 logit to 1.42 logit, indicating a substantially higher ADHD scoring for Chinese children. Additionally, the analysis of symptom ratings indicated that Chinese children received significantly higher scores than English children in 'Inattentive' and 'Distracted' symptoms.

The study also explored the impact of anchoring teacher leniency on gender differences and found that rater leniency is consistent across girls and boys. In both countries, girls received significantly lower ratings on ADHD compared to boys. After anchoring the teacher's leniency, both Chinese girls and boys showed a significant increase in their measures, while the difference between girls and boys remained unchanged. Similarly, after anchoring the teachers' leniency, both English girls' and boys' measures decreased significantly, with no significant change in the gender difference. A Two-way ANOVA analysis revealed no interaction effect of gender and country factors on children's measures.

The gender difference in the study aligns with other research findings, where boys' ADHD measures are significantly higher than girls' (Polanczyk et al., 2007; Gershon, 2002). Contrary to Nussbaum's suggestion (2012) that girls with ADHD may face more difficulties with inattentive symptoms than hyperactive and impulsive symptoms, the current study found that girls were rated significantly lower than boys on all inattention, hyperactivity, and impulsivity symptoms in both countries. Schultz and Evans (2012) conducted a similar investigation on rater leniency's effect on adolescents' ADHD behaviours and reported consistent leniency patterns across genders, which aligns with the findings of the present study. This consistency in rater leniency across genders emphasizes the importance of considering and controlling for potential bias when assessing children's ADHD behaviours.

Upon adjusting children's measures with their raters' leniency, significant changes in the children's ratings were observed, leading to an increased difference between the two countries. This finding underscores the substantial impact of raters' leniency differences on the assessment of children's behaviour. While a limited number of cross-cultural studies have

explored differences in how teachers perceive and rate children's ADHD behaviours between China and the UK, none have quantified the extent to which these differences affect the ratings of children's ADHD behaviour (Du et al., 2003; Alban-Metcalfe et al., 2002).

Conducting cross-rating to quantify rater leniency is crucial, yet it poses challenges in terms of time and expenses for cross-country studies. An alternative approach is to develop an instrument capable of measuring teachers' leniency across different cultures, while ensuring it employs appropriate media, such as video tape or cartoon animation vignettes, to avoid introducing confusion and potential biases. In this study, a series of cartoon animations were carefully designed to portray children's ADHD behaviours for raters to observe and rate to effectively measure raters' leniency. Leveraging the Many-facet Rasch Model, the study successfully adjusted children's ADHD behaviours with measured rater leniency. Thus, this study stands as the pioneering investigation into the impact of rater leniency on children's ADHD behaviour ratings across both England and China.

The study reveals a significant disparity in measures between Chinese and English children, with Chinese children's measures notably higher. This observation implies that the estimated prevalence of children with ADHD in China might surpass that in England. Remarkably, this finding aligns with prevalence studies conducted in both countries. In the UK, a few studies have estimated the prevalence of ADHD in children to be around 1.5% to 2% (Russell et al., 2014; Green et al., 2005). In contrast, meta-analysis studies in China have reported a higher estimated prevalence ranging from 5.6% to 6.26% (Lian et al., 2013; Wang et al., 2017). These findings collectively suggest a higher prevalence of ADHD among Chinese children compared to their English counterparts, which aligns with the present study.

However, findings from the prevalence studies conducted in China and UK are not comparable. In the UK, estimated prevalence figures are derived from national survey results, providing a standardised and free diagnostic procedure within the NHS system. Skilled paediatricians evaluate information from parents and teachers to ascertain the number and severity of symptoms, along with related impairments. However, in China, the lack of medical professionals and expensive medical diagnostic costs, studies did not adopt such standardised procedures for diagnosing ADHD (Leadleo, 2021). Consequently, studies conducted in various regions of China often rely on surveys to collect information from teachers, parents, or both to estimate prevalence, without the involvement of professional clinicians. Some studies even employ a simple rule for issuing a positive diagnosis of ADHD, such as reaching a particular cut-off score in both parent and teacher ratings. This approach may result in a relatively higher prevalence with potential for false positives. Therefore, the disparity in estimated ADHD prevalence between China and the UK does not necessarily imply a higher number of children with ADHD in China than in the UK. Instead, it may reflect the impact of different diagnostic procedures and study methodologies in each country.

10.5 Research Question 4

To what extent are the ratings of ADHD behaviours comparable from teacher to teacher in the Chinese sample?

As we mentioned in the previous chapter of the Literature Review, teacher ratings of behaviour are the most used method in the school-based assessment of ADHD (Barkley, 2006). The school is a more demanding environment than home, where problematic behaviours become more pronounced. Meanwhile, the new diagnostic criteria of DSM-V and ICD-11 both request multi-informants across settings about the manifestation of symptoms and impairments of ADHD. Therefore, schools have become an essential setting for ADHD assessment. At school, teachers have the opportunity to compare children in a large cohort and a wide range of activities. Their professional training and teaching experience help them to recognize

problematic behaviours. Therefore, teachers play an important role in the diagnosis of ADHD. Although the new diagnostic criteria request multi-informants across different settings to reach a level of agreement to ensure a positive diagnosis, it does not require multi-informants for each setting. Research indicates that agreement among multi-informants tends to be low to moderate (Achenbach et al., 1987; Murray et al., 2007). Since my study is focused on presenting a fair comparison of teachers' ratings of children's ADHD ratings between China and England, I was also interested in how reliable it is to rely on one teacher's judgement about a child's ADHD behaviours.

Due to the teaching structure in China, it is relatively easy to collect ratings from multiple teachers for a child. I have conducted multi-informants' ratings for 400 Chinese children in 40 groups selected randomly from Year 2 classes in China. Each group of children were rated by a team of 2 to 5 teachers about the behavioural frequency of six ADHD symptoms (Can't Wait, Fidgety, Inattentive, Disorganised, Distracted, Disruptive). Then I conducted an Intraclass Correlation Coefficient (ICC) analysis to investigate the absolute agreement of each team of teachers' ratings. I found that less than half of the ADHD ratings are comparable from teacher to teacher in China. It is unreliable to rely on a single teacher's judgement about a child's ADHD symptoms. Although the average agreement of the 40 groups of teachers is fair, the agreements vary in a wide range. About 22.5% of the teachers either have an opposite opinion or a low agreement about their ratings. 35% of teachers have reached a fair level of agreement with their colleagues. However, 49% to 60% of unexplained variances are in their ratings. Less than half (42.5%) of teachers have good to excellent (.60 to .89) levels of agreement with each other. These teachers' ratings are comparable with each other.

My study finding of the interrater agreement between Chinese teachers is relatively lower than in some studies. Loughran (2003) has conducted longitudinal research to investigate the ratings of teachers and teacher assistants of children's ADHD behaviours in a cohort of preschool children at two different times. She found that the agreement between teachers and teaching assistants at the two different times is .61 and .75, respectively. It is relatively high compared to my study findings. However, her research only involved 60 children, 3 teachers and 3 teaching assistants. These children, teachers and teaching assistants were all from the same school.

Achenbach et al. (1987) carried out a meta-analysis of 119 studies which involved 269 samples aged from 1 to 19 years old, to investigate the degree of consistency between different informants when they report behavioural and emotional problems. They found the mean correlation of paired teachers' agreement is .64 (1987). However, the 20 teacher-to-teacher correlations they examined from different studies covered a wide range of children from 18 months to 16 years old. It covers a wide range of settings from nursery to secondary schools. Therefore, ratings were made in different situations. Some were made by teachers and teaching assistants who worked with the same group of children in the same classroom. While others were made either by team teachers or teachers who taught in different classes. Moreover, the studies included in the analysis are mainly from the USA.

There are other studies which have found slightly lower interrater agreement. Molina et al. (1998) found interrater agreements about adolescents' ADHD behaviours with average ICC estimates ranging between .46 to .48 among secondary school teachers. In their study, each adolescent has 2 to 5 teachers rate his behaviour with a broadband behavioural rating list. However, their study only involved male adolescents who had a childhood history of ADHD.

Schultz and Evans (2012) investigated the ICC estimates of secondary teachers about adolescents' ADHD behaviours with school samples. They found that the ICC estimates range between .46 to .55.

My study differs from theirs in the children's age group, cross-rating type and statistics analysis. The children in my study are in the same age group (Year 2). The cross-informant ratings are complete cross-rating. The 40 groups of teachers (N=89) are from 11 schools, and each group of teachers all rated the same 10 randomly selected children from their class. The interrater agreements were calculated with individual groups of children and matched group teachers on individual symptoms' ratings. I have used the Intraclass Correlation Coefficient with the two-way-mixed model to calculate the absolute agreement and recorded the average measures of individual groups of children on individual items.

I have conducted the study and analysed data in this way because I am more interested in the absolute agreement among teachers of individual groups rather than a general reliability index for the whole scale. Since I need to pick only a single teacher's rating from each group of teachers for later analysis, this gives me a more detailed indication of how reliable the rating I would pick from each group. For example, a group of two teachers (t052, t053) has a negative average ICC estimate -.12 for 10 children they rated. The individual symptoms ICC estimates show us they have either 0 or negative agreement on all six symptoms. The most likely reason for that kind of disagreement between them is that one of the teachers might oppositely use the categories. Instead of rate 0 for never and 4 for always, one of the teachers might rate 4 for never and 0 for always. Another group of low agreement (.17) teachers (t010, t011) have reached a good level of agreement for the symptom 'Inattentive' (.70). However, they disagreed with each other on symptoms of 'Can't wait (-.46) and 'Disruptive' (-.11). It indicates that they

either interpreted these two symptoms in a completely different way or they have completely different tolerance to these two symptoms.

Meanwhile, the average ICC estimates of individual symptoms show us to what extent the 40 groups of teachers' ratings agree with each other on individual symptoms. I found that teachers have different levels of agreement with individual symptoms. The average ICC estimates from 40 groups of teachers on the six symptoms varied between .47 to .64. Teachers' ratings have reached a fair level of agreement on five symptoms (Can't wait, Fidgety, Disorganised, Distracted and Disruptive) with ICC ranging from .47 to .55. It indicates that teachers have a fair level agreement with their colleagues on the interpretation and tolerance of these symptoms when they rate the same group of children from their class. One symptom 'Inattentive' has a good value of ICC of .64. It indicates that teachers have a good level of agreement with their colleagues on the interpretation and tolerance of inattentive symptoms when they rate the same group of children in their class. This symptom is well recognised by teachers and could be a very important indicator for predicting children's future academic achievement.

Other studies have similar findings. Sollie et al. (2013) compared mother and father's ratings of ADHD symptoms; they found that the ICC estimates between mother and father of the individual symptoms varied between .41 and .77. However, they only reported the range of ICC estimates for all 18 ADHD symptoms rather than individual symptoms. My findings of the ICC estimates of the six symptoms are within their range. A more recent study has reported similar findings about the interrater agreement between mother and father about the six symptoms of ADHD symptoms. They reported the ICC estimates of individual symptoms. They found that five symptoms' (Can't wait, Fidgety, Distracted, Disruptive, Inattentive) ICC estimates are in the range of .50 to .60 (Caye et al., 2017). However, they found the symptom

of ‘Disorganised’ with a relatively low ICC estimate (.18). Another study also found a very low interrater agreement between parents on ‘Disorganised’ (-.04) while the other five symptoms (Can’t wait, Fidgety, Distracted, Disruptive, Inattentive) are in a range of .31 to .65 (Martel et al., 2015). Meanwhile, the symptom Disorganised has the lowest ICC estimate (.47) among the six symptoms in my study.

There are possible reasons for the relatively low interrater agreement on symptom ‘Disorganised’ compared to other symptoms. One reason might be because this symptom is not well defined in the diagnostic description therefore informants have different interpretations of what behaviours show disorganised symptoms. Therefore, even in the same environment for the same children, agreements among different informants are lower on disorganised symptoms compared to other symptoms.

When teachers have a high degree of agreement, it indicates that the independent rater has only introduced a minimal amount of measurement error and the statistical power for subsequent analysis is not reduced. When teachers have a low degree of agreement when they rate the same child on the same behaviours. Many factors could contribute to the discrepancies among raters, systematic differences related to the informants (cultural expectation), other child's psychopathology and situated specificity of children's behaviours (Dirks et al., 2012). The symptoms of diagnostic criteria (DSM-V or ICD-10) are only general descriptions without context details. They leave space for raters to interpret what behaviour and frequency these behaviours fall into the category of ADHD in a specific context. Individual teachers have different thresholds and personal standards when rating problem behaviour (Van der Ende, 1999). These standards depend on an individual's life experience and knowledge of typical child development, and expectations towards children. Therefore, teachers may have different

interpretations of the behaviours, such as behaviour 'disorganised'. Some teachers may interpret a messy table as disorganised, while others may interpret unpredictable agitation as disorganised. Meanwhile, teachers may have different interpretations of the rating categories. For example, some teachers may think 1 or 2 times a day is qualified as 'Occasionally' while others may think it is 'Very Often'. It could also be due to the different tolerance of individual teachers to particular behaviours.

However, findings are often difficult to compare directly because different studies have different sample characteristics, informant characteristics, assessment methods, procedures and statistical analyses. All these factors may influence research findings. The sample characteristics may contribute to the discrepancy between research findings (Amador-Campos et al., 2006, Giretens et al., 2004). Some studies recruited children from clinically referred patients while others recruited from the community population. Research shows that agreement between informants of children's symptoms is generally higher for clinically referred children than for community samples. Meanwhile, children's age may contribute to the discrepancy between informants. Some studies found the agreement between mother and father is higher for older children than for younger children while other studies have contradictory findings (Achen-bach et al. 1987; Duhig et al.2000).

The type of informants may contribute to the discrepancy between them. Studies show that the interrater agreement of different informants is different (O'Neill et al., 2013; Sollie et al., 2013). One study found that the interrater agreement between parents and clinics is higher than the agreement between teachers and clinics (O'Neill et al., 2013). At the same time, Sollie et al. (2013) found that the agreement between father and teachers is higher than the agreement between mother and teachers. A meta-analysis study of 269 samples found that the agreement

of informants from the same setting (e.g. mother and father/ Paris of teachers) is higher than the agreement of informants from different settings (e.g. parents and teacher) (Achenbach et al., 1987). Different types of informants, such as mothers, fathers, teachers or clinical professionals, may have different perceptions of ADHD symptoms.

Furthermore, they observe children's behaviours in different environments and interpret children's behavioural problems in different contexts. Meanwhile, a child could behave differently in a different environment. The informants' education level, mental health situation and the role they play in the child's life may also influence their ratings of children's ADHD symptoms (Schults & Evans,2012; Caye et al., 2017). Therefore, the type of informants the studies used to investigate may potentially influence their findings.

It is the same for assessment methods and measurements. Studies have used various methods to collect ratings from informants, such as standardised questionnaires, structured interviews, and observation. Different methods have their advantages and unique errors. Some researchers have generated interrater agreement between two groups of informants with two different assessment methods. For example, they used questionnaires to collect teachers' ratings and interviews to collect parents' ratings. The difference between the two assessment methods could contribute to the discrepancy between the two rating groups. Some studies used measurements for broadband measurements such as CBCL (Children's Behaviour Checking List) to collect informants' ratings about not only ADHD symptoms but also children's behavioural and emotional problems. In contrast, other studies used relatively narrow measurements to collect informants' ratings about 18 ADHD symptoms. Different measurements have different ways of presenting the rating scheme and describing the behaviours. They could potentially influence the study findings.

Apart from the factors discussed above, statistics play an important role in influencing the research findings. Studies have adopted a wide range of statistics and standards to generate Interrater agreement estimates and thresholds. First, there are many different statistics for interrater agreement, such as Choehn's kappa, Fleiss' kappa, Weighted kappa, Kendall coefficient of concordance, Bland-Altman plots and Intraclass Correlation Coefficient (ICC). Some studies applied traditional statistics to generate it, such as Pearson correlation and percentage agreement. However, their limitations are obvious. Pearson correlation only reflects the association of paired raters but does not reflect the actual disagreement in rating levels (Bartko, Carpenter, 1976). Percentage agreement does not take into account chance agreement that has a big impact on rater's rating when they rate on short categories. Although different researchers presumably selected the appropriate statistics according to their study purpose and research design, each statistic has its limitations regarding the number of raters and data type of the measurement that the statistic can process (Gisev et al., 2013).

Second, there are two types of agreements for interrater agreements. One type is the absolute agreement which relates to which different raters assign the same precise value for each item being rated. Another type is consistency which relates to the extent to which raters can consistently distinguish between different items on a measurement scale (Landers, 2015). The absolute agreement is sensitive to the differences in leniency in ratings, while consistency is sensitive to the ordering of ratings. A high value of consistency could have a low absolute agreement because the two types of statistics calculate different types of agreements. Not all the studies report what kind of agreement their analysis carried out.

Third, there is no universal agreement as to the acceptable value of the interrater agreement. Although each statistic has made standards for different levels of agreement (poor, fair, good, excellent), they are not comparable. For example, Kappa standards interpret .21 to .40 as a fair level of agreement, while with ICC standards, it was interpreted as a poor level of agreement (Shweta et al., 2015). Even with the same statistic, different standards studies have applied, such as the statistic ICC I have applied. There are three standards that different studies have adopted, shown in *Table 90*.

Table 90 ICC standards

References	Standards
(Koo & Li, 2016)	Less than .5 poor Between .5 and .75 moderate Between.75 and .9 good Greater than .9 excellent
(O'Neill, 2017)	Less than .3 Lack of agreement Between .31 to .50 week Between .51 to .70 moderate Between .71 to .90 strong Between .91 to 1 very strong
(Cicchetti, 1994)	Less than .40 Poor Between .40 and .59 Fair Between .60 and .74 Good Between .75 to 1 Excellent

It is not comparable if two researchers performed different statistics and standards for the interrater agreement. However, even using the same statistic, such as ICC, different models generate different estimates. In general, the absolute agreement gives a smaller ICC estimate than the consistency, and the estimate of the single measurement is lower than the average measurements estimate in the ICC estimate (Koo & Li, 2016). Some studies did not report the model details of ICC they performed for their data (Sollie et al., 2013; Caye et al., 2017). Due to the reasons discussed above, it is hard to compare studies directly.

I was lucky that I was able to conduct my research with 40 groups of teachers across 11 schools to investigate the comparability of children's ADHD ratings with multi-informants in China. One advantage of conducting this type of research in China is that each Year 2 class was taught by 5 to 6 subject teachers. This teaching structure helps me to collect each child's symptoms ratings from 2 to 5 teachers. However, each teacher teaches more than one class due to the large pupil population at schools in China. Different subjects have different weights in children's learning. Subjects such as PE or Crafts only have one or two weekly sessions for the pupils.

Meanwhile, the classes of the schools I have access to are from 30 to 60 pupils in one classroom. It indicates that some Chinese subject teachers may not know the pupils well due to the limited time and tension they can spend on each child. Unlike in China, the Year 2 class in England is normally taught by one teacher and a teaching assistant (if available). They teach all the subjects and spend all day with the children. Meanwhile, the class size is smaller in England than in China. It will be interesting to investigate the interrater agreement between teachers and teaching assistants in England and explore what factors impact teachers' ratings in both countries in future research.

10.6 Conclusion

In conclusion, this chapter's discussion sheds light on some possible advancements made in the assessment of children's ADHD behaviours in cross-cultural contexts. By developing and employing the innovative Teacher's Rating Scale of Cartoons (TRSC) and utilising the Many-Facet Rasch Model (MFRM) with anchored leniency measures, the study has addressed the pervasive issue of rater leniency, enhancing the fairness and accuracy of cross-cultural comparisons. The study's findings contribute significantly to the field of cross-cultural assessment, offering valuable insights into the role of cultural and contextual factors in shaping

teachers' perceptions of children's ADHD behaviours. Moreover, the successful application of the TRSCB and TRSC across Chinese and English samples underscores its cross-cultural applicability and potential for standardised use in various educational and psychological evaluations. These insights hold promising implications for fostering a more inclusive and equitable assessment environment for children with ADHD in diverse cultural settings. Moving forward, future research endeavours should explore additional factors influencing teachers' ratings and validate the generalisability of these findings to other cultural contexts. Ultimately, this research represents a significant step forward in promoting fairness, accuracy, and sensitivity in the assessment of children's ADHD behaviours, ultimately supporting improved understanding and support for children with ADHD in educational and clinical settings.

Chapter 11 Conclusion

This study represents a pioneering effort in developing an innovative rating scale that utilises cartoon animations to measure teachers' leniency towards children's ADHD behaviours in cross-cultural contexts. Furthermore, to ensure fair and accurate cross-cultural comparisons, the researcher employed the Many-Facet Rasch Model (MFRM) to address the bias of rater leniency, which is a common issue in rating scales. However, due to practical constraints, cross-rating was not a feasible option for this cross-country study. Instead, the researchers innovatively utilised anchoring techniques, separately measuring teachers' leniency through the Teacher's Rating Scale of Cartoons (TRSC) and introducing these measures as fixed facets in the MFRM analysis of TRSCB. The results of the MFRM analysis with anchored leniency measures successfully adjusted children's ADHD behaviour measures, accounting for variations in teachers' leniency. This approach ensured that the measures of children's ADHD behaviours were fair and consistent across different cultural contexts, allowing for meaningful cross-cultural comparisons. The findings of this study have several important implications for research, education, and clinical practice in the field of ADHD behaviours and cross-cultural studies.

Implications

The use of cartoon animations as rating objects in measuring teachers' leniency towards ADHD behaviours offers a more culturally sensitive approach compared to traditional methods like written vignettes or video recordings. By presenting behaviours in visual and engaging formats, this approach accounts for potential cultural biases introduced by language and translation, making it suitable for cross-cultural measurement. Researchers and educators can consider adopting similar innovative assessment methods to improve the validity and fairness of evaluations in diverse cultural contexts. The development and validation of measuring instruments TRSC and TRSCB demonstrated the significance of using culturally sensitive

approaches to develop instruments for comparing ADHD behaviours between different countries. These scales can be adapted and used in various educational and clinical settings to ensure that assessments of ADHD behaviours are contextually relevant and conducted with minimised bias across different cultural contexts. Future studies should consider cultural nuances in symptom expression and incorporate culturally appropriate stimuli to enhance the validity of ADHD assessments across diverse populations.

This study demonstrated that cartoon animation is an effective way to portray children's ADHD behaviours as observing objects to explore rater bias. Expanding on this approach, future studies could utilise cartoon animations to delve into biases that may exist within the home setting, involving both mothers and fathers as informants. The versatility of cartoon animations lies in their potential to be developed with minimal cultural bias, making them applicable across diverse cultures, languages, and even within a single culture. For instance, these animations can be employed to compare how teachers from different regions within the UK, such as Scotland, Wales, and England, rate ADHD behaviours. Similarly, they could shed light on potential differences in ratings between Chinese teachers from urban and rural areas. Thus, the cartoon animation approach emerges as a valuable tool to explore rater bias both within and across cultures. This study serves as a trial study for future further studies.

The use of the Many-Facet Rasch Model (MFRM) with anchored leniency measures offers a novel approach to address the issue of rater leniency bias in cross-cultural studies. This technique allows for fair and accurate comparisons of ADHD behaviours between different cultural groups, ensuring that cultural differences in teachers' perceptions do not confound the results. The study's findings underscore the importance of considering and controlling for rater bias in cross-cultural studies, especially when assessing sensitive and complex behaviours like

ADHD. Rater bias can significantly impact the interpretation and comparison of ADHD prevalence and severity between different cultural contexts. Researchers conducting cross-cultural studies on ADHD should be mindful of potential rater leniency effects and employ appropriate statistical techniques, such as the MFRM with anchoring, to adjust for bias and ensure fair and accurate assessments. The significant differences in how Chinese and English teachers rated certain ADHD behaviours also highlight the influence of cultural factors on teachers' attitudes. These cultural variations should be carefully considered when conducting cross-cultural studies on ADHD behaviours and when developing interventions and support strategies for children with ADHD in different cultural settings. Understanding these differences can inform the development of culturally relevant and context-specific approaches to address ADHD-related challenges. Future studies, such as teacher interviews and surveys are needed to further investigate the reasons that may cause the differences in how teachers perceive ADHD behaviours between the two countries.

The substantial difference in children's ADHD measures between China and England suggests that estimating ADHD prevalence may not be directly comparable due to variations in diagnostic procedures, methodologies and rater bias. Clinicians and researchers should be cautious when making direct comparisons between ADHD prevalence rates from different cultural contexts. This highlights the need for standardised diagnostic procedures, consistent methodologies and culturally sensitive approach when estimating ADHD prevalence globally. The study's findings call for the development of standardised and culturally sensitive diagnostic procedures for ADHD. Furthermore, more cross-cultural studies are needed in ADHD study because cross-cultural studies play a vital role in advancing our knowledge of ADHD by highlighting cultural influences on symptom presentation, diagnosis, and treatment. By addressing the complex interplay between cultural factors and ADHD, these studies can inform

the development of culturally sensitive and effective interventions, leading to improved outcomes for children and individuals with ADHD across diverse cultural contexts.

The findings regarding the influence of teaching experience on teachers' leniency towards ADHD behaviours underscore the importance of targeted training and professional development for educators. Educating teachers about ADHD symptoms and diagnostic criteria, can enhance their understanding and judgement when rating children's behaviours. Teachers with different levels of experience and from diverse cultural backgrounds may benefit from specific interventions to enhance their awareness and understanding of ADHD behaviours and develop effective classroom management strategies.

This study also delved into the reliability of using one teacher's judgement in China, where collecting ratings from multiple teachers for a child is relatively easier due to the teaching structure. The results of the study revealed that relying on a single teacher's judgement about a child's ADHD symptoms may not be reliable in China. The inter-rater agreement investigation in the Chinese sample highlights the significance of collecting ratings from multiple informants in the same setting to obtain a reliable assessment of children's ADHD behaviours. Relying on a single teacher's judgement may not be sufficient to capture accurate behaviours exhibited by a child in the school setting. Although the diagnostic criteria do not require multiple informants in the same setting, including multiple teachers' perspectives can lead to more accurate and reliable assessments. The study also highlights variations in interrater agreement for different ADHD symptoms. The symptoms that demonstrate low agreement among raters may require further clarification and detailed descriptions in diagnostic criteria to ensure consistent interpretations across different settings and cultures.

The study's findings may have implications for designing targeted intervention strategies for children with ADHD. Understanding specific symptom patterns and variations between countries can help tailor interventions to address the unique needs of children in different cultural contexts. For example, this study findings indicate that Chinese children manifest higher levels of inattentive and distracted behaviours than English children in the classroom. It may indicate that the Chinese teaching style (teacher-centred lecturing) may cause children to lose interests and attention easily. Interventions including diverse learning activities, such as collaborative learning, games and quizzes, and role playing may help children with inattentive and distracted behaviours maintain focus and achieve their learning goals.

Limitations

Limited Sample Size and Sampling Method: The sample size in this study, particularly for teachers from England (N=24), was relatively small. Meanwhile, the study used a convenience sampling approach not random sampling. The limited sample size and sampling approach may affect the generalizability of the findings and may not fully represent the diversity of perspectives within the two cultural contexts. Future research should strive to include larger and more diverse samples to strengthen the reliability and representativeness of the results.

Restricted ADHD Behaviours: Although the study developed a behavioural bank for all ADHD symptoms, with limited time and budget, there is only a series of six episodes of cartoon animations produced to portray children’s behaviours in six symptoms. Therefore, the study focused on a specific set of ADHD symptoms in the TRSC and TRSCB measurements, namely "Can't wait," "Fidgety," "Inattention," "Disorganised," "Distracted," and "Disruptive". While these symptoms were carefully selected and validated, they may not capture the entire spectrum of ADHD-related behaviours. ADHD is a complex and multifaceted disorder with various symptoms. Future studies should consider incorporating a more comprehensive set of ADHD

behaviours and develop more cartoon animations to capture the full complexity of the construct to provide a more nuanced understanding of ADHD across different cultural backgrounds.

Teacher Bias: This study has shed light on the significant differences in how teachers from China and England rated the portrayed ADHD behaviours. However, the underlying reasons behind these differences remain unknown. Conducting follow-up studies, such as teacher interviews or surveys, could provide valuable insights into the factors influencing these variations. It is also important to note that this study primarily focused on rater leniency as one type of rater bias affecting the ratings. Nevertheless, it is essential to recognize that rater bias encompasses various forms, including the "halo" effect and central tendency. Future research should delve into investigating these other types of rater bias to gain a comprehensive understanding of their impact on ratings.

Potential Bias in Animation Production: While efforts were made to ensure cultural neutrality in the animations, some cultural biases may still be present. Future research should engage diverse cultural perspectives and validate the animations in different cultural contexts to minimise potential bias for further applications.

ADHD Prevalence Comparisons: While the study revealed a significant difference in ADHD measures between Chinese and English children, caution should be exercised when interpreting this as an indication of a higher prevalence of ADHD in China than in England. It is important to note that the study solely focused on six symptoms, representing only one third of the full diagnostic criteria. Consequently, the measurement was limited in capturing the full spectrum of ADHD symptoms. As a result, drawing definitive conclusions regarding the actual differences in ADHD behaviours between children from China and England would be

challenging. Further research with a more comprehensive approach is warranted to gain a more accurate understanding of ADHD prevalence in both regions.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232. https://doi.org/10.1037/0033-2909.101.2.213

AI-Hindawe, J. (1996). Considerations when constructing a semantic differential scal. *School of Communication, Arts and Critical Enquiry*, (1036–0808), 41–58. https://doi.org/http://hdl.handle.net/1959.9/509873

Alban-Metcalfe, J., Cheng-Lai, A., & Ma, T. (2002). Teacher and Student Teacher Ratings of Attention-Deficit/Hyperactivity Disorder in Three Cultural Settings. *International Journal of Disability, Development and Education*, *49*(3), 281–299. https://doi.org/10.1080/1034912022000007298

Albaum, G., Best, R., & Hawkins, D. (1981). Continuous vs Discrete Semantic Differential Rating Scales. *Psychological Reports*, *49*(1), 83–86. https://doi.org/10.2466/pr0.1981.49.1.83

Alexiou, T. (2019). Cartoons that Make A Difference: A Linguistic Analysis of Peppa Pig. *Journal of Linguistics and Education Research*, *1*(1). https://doi.org/10.30564/jler.v1i1.314

Altemeier, W. A., & Horwitz, E. (1997). The Role of the School in the Management of Attention Deficit Hyperactivity Disorder. *Pediatric Annals*, *26*(12), 737–744. https://doi.org/10.3928/0090-4481-19971201-13

Anastopoulos, A. D., & Shelton, T. L. (2001). *Assessing attention-deficit/hyperactivity disorder* (1st ed.). Kluwer Academic/Plenum Publishers.

Anderson, J. C. (1996). Is childhood hyperactivity the product of western culture? *The Lancet*, *348*(9020), 73–74. https://doi.org/10.1016/s0140-6736(05)64598-9

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 357–374. https://doi.org/10.1007/bf02293814

Aramwit, P., & Wirotsaengthong, S. (2013). Overview of commonly used Chinese herbs. *Journal of Medicinal Plants Research*, *7*(16), 998–1014. https://doi.org/10.5897/jmpr11.481

Arlin, M. N., & Hills, D. (1974). Comparison of Cartoon and Verbal Methods of School Attitude Assessment Through Multitrait-Multimethod Validation. *Educational and Psychological Measurement*, *34*(4), 989–995. https://doi.org/10.1177/001316447403400434

Arneklev, B. J., Cochran, J. K., & Gainey, R. R. (1998). Testing Gottfredson and Hirschi's "low self-control" stability hypothesis: An exploratory study. *American Journal of Criminal Justice*, *23*(1), 107–127. https://doi.org/10.1007/bf02887286

Arnold, L. E. (1996). Sex differences in ADHD: Conference summary. *Journal of Abnormal Child Psychology*, *24*(5), 555–569. https://doi.org/10.1007/bf01670100

Arthur, J., Waring, M., Coe, R., & Hedges, L. V. (2017). *Research methods and methodologies in education* (1st ed.). SAGE PUBLICATIONS.

Aryadoust, V. (2016). Gender and academic major bias in peer assessment of oral presentations. *Language Assessment Quarterly*, *13*(1), 1–24. https://doi.org/10.1080/15434303.2015.1133626

Aryadoust, V., & Raquel, M. (2019). *Quantitative data analysis for language assessment Volume I: Fundamental Techniques* (1st ed.). Routledge.

Bader, M., & Hadjikhani, N. (2014). The concept of instability: a French perspective on the concept of ADHD. *ADHD Attention Deficit and Hyperactivity Disorders*, *6*(1), 11–17. https://doi.org/10.1007/s12402-013-0121-y

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., Hoza, B., Hoza, B., Pelham, W. E., Molina, B. S. G., & Bagwell, C. L. (2001). Attention-Deficit Hyperactivity Disorder and Problems in Peer Relations: Predictions From Childhood to Adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1285–1292. https://doi.org/10.1097/00004583-200111000-00008

Barkley, R.A. (2002). Major life activity and health outcomes associated with attention-deficit/hyperactivity disorder. *The Journal of Clinical Psychiatry.*, *12*(65), 10–15.

Barkley, Russell A, & Murphy, K. R. (1998). *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis andTreatment [and] a Clinical Workbook. Second Edition* (1st ed.). Guilford Press.

Barkley, Russell A. (2012). Distinguishing sluggish cognitive tempo from attention-deficit/hyperactivity disorder in adults. *Journal of Abnormal Psychology*, *121*(4), 978–990. https://doi.org/10.1037/a0023961

Barkley, Russell A. (2013). Distinguishing Sluggish Cognitive Tempo From ADHD in Children and Adolescents: Executive Functioning, Impairment, and Comorbidity. *Journal of Clinical Child & Adolescent Psychology*, *42*(2), 161–173. https://doi.org/10.1080/15374416.2012.734259

Barkley, Russell A., & Peters, H. (2012). The Earliest Reference to ADHD in the Medical Literature? Melchior Adam Weikard's Description in 1775 of "Attention Deficit" (Mangel der Aufmerksamkeit, Attentio Volubilis). *Journal of Attention Disorders*, *16*(8), 623–630. https://doi.org/10.1177/1087054711432309

Barkley, Russell A., Fischer, M., Smallish, L., & Fletcher, K. (2002). The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *Journal of Abnormal Psychology*, *111*(2), 279–289. https://doi.org/10.1037//0021-843x.111.2.279

Barkley, Russella A, Biederman, J., Biederman, J., & Barkley, R. A. (1997). Toward a Broader Definition of the Age-of-Onset Criterion for Attention-Deficit Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(9), 1204–1210. https://doi.org/10.1097/00004583-199709000-00012

Batchelor, B. (2009). *American pop*. Greenwood Press.

Bauermeister, J. J., Shrout, P. E., Ramírez, R., Bravo, M., Alegría, M., Martínez-Taboas, A., Chávez, L., Rubio-Stipec, M., García, P., Ribera, J. C., & Canino, G. (2007). ADHD Correlates, Comorbidity, and Impairment in Community and Treated Samples of Children and Adolescents. *Journal of Abnormal Child Psychology*, *35*(6), 883–898. https://doi.org/10.1007/s10802-007-9141-4

Baughman Jr., F. A. (2006). *The ADHD Fraud: How Psychiatry Makes 'Patients' of Normal Children* (1st ed.). Trafford Publishing.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control* (1st ed.). Academic Press.

Beau-Lejdstrom, R., Douglas, I., Evans, S. J. W., & Smeeth, L. (2016). Latest trends in ADHD drug prescribing patterns in children in the UK: prevalence, incidence and persistence. *BMJ Open*, *6*(6), e010508. https://doi.org/10.1136/bmjopen-2015-010508

Benton, J. R. (2009). *Materials, methods, and masterpieces of medieval art*. Praeger.

Bergey, M. R., Filipe, A. M., Conrad, P., & Singh, I. (2018). *Global perspectives on ADHD: Social dimensions of diagnosis and treatment in sixteen countries.* (1st ed.). Johns Hopkins University Press.

Biederman, J., Faraone, S. V., Weber, W., Russell, R. L., Rater, M., Park, K. S., Park, K. S., Rater, M., Russell, R. L., Weber, W., Faraone, S. V., & Biederman, J. (1997). Correspondence Between DSM-III-R and DSM-IV Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(12), 1682–1687. https://doi.org/10.1097/00004583-199712000-00016

Biederman, J., Wilens, T. E., Mick, E., Faraone, S. V., & Spencer, T. (1998). Does attention-deficit hyperactivity disorder impact the developmental course of drug and alcohol abuse and dependence? *Biological Psychiatry*, *44*(4), 269–273. https://doi.org/10.1016/s0006-3223(97)00406-x

Birnbaum, H. G., Kessler, R. C., Lowe, S. W., Secnik, K., Greenberg, P. E., Leong, S. A., & Swensen, A. R. (2005). Costs of attention deficit–hyperactivity disorder (ADHD) in the

US: excess costs of persons with ADHD and their family members in 2000. *Current Medical Research and Opinion*, *21*(2), 195–205. https://doi.org/10.1185/030079904x20303

Bishop, D. V. M. (2010). Which Neurodevelopmental Disorders Get Researched and Why? *PLoS ONE*, *5*(11), e15112. https://doi.org/10.1371/journal.pone.0015112

Bloom, B., Jones, L. I., & Freeman, G. (2013). *Summary Health Statistics for U.S. Children: National Health Interview Survey, 2012* (Ser. Vital and Health Statistics Series 10, Number 258, p. 4). Washington DC: US Department of Health and Human Services.

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2005). *Handbook of self-regulation* (1st ed.). Academic Press.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.). LAWRENCE ERLBAUM Associates (NJ).

Borg , M. G., Falzon, J. M., Falzon, J. M., & Borg , M. G. (1990). Teacher's perception of primary school children's undesirable behaviours: the effects of teaching experience, pupil's age, sex and ability stream. *British Journal of Educational Psychology*, *60*(2), 220–226. https://doi.org/10.1111/j.2044-8279.1990.tb00939.x

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America* (1st ed.). Basic Books.

Brewis, A., Schmidt, K. L., & Meyer, M. (2000). ADHD-Type Behavior and Harmful Dysfunction in Childhood: A Cross-Cultural Model. *American Anthropologist*, *102*(4), 823–828. https://doi.org/10.1525/aa.2000.102.4.823

Brown, R. T., Amler, R. W., Freeman, W. S., Perrin, J. M., Stein, M. T., & Feldman, H. M. (2005). Treatment of Attention-Deficit/Hyperactivity Disorder: Overview of the Evidence. *PEDIATRICS*, *115*(6), 749–757. https://doi.org/10.1542/peds.2004-2560

Bujang, M. A., & Baharum, N. (2017). A simplified guid to determine of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *The Journal of the School of Dental Sciences*, *12*(1), 1–11.

Cambridge University Press. (2019). Cartoon. *Cambridge Dictionary*.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation* (1st ed.). Springer.

Caye, A., Machado, J. D., & Rohde, L. A. (2017). Evaluating Parental Disagreement in ADHD Diagnosis: Can We Rely on a Single Report From Home? *Journal of Attention Disorders*, *21*(7), 561–566. https://doi.org/10.1177/1087054713504134

Chapple, C. L., Vaske, J., & Hope, T. L. (2010). Sex differences in the causes of self-control: An examination of mediation, moderation, and gendered etiologies. *Journal of Criminal Justice*, *38*(6), 1122–1131. https://doi.org/10.1016/j.jcrimjus.2010.08.004

Chin, C., & Teou, L.-Y. (2010). Formative assessment: Using concept cartoon, pupils' drawings, and group discussions to tackle children's ideas about biological inheritance. *Journal of Biological Education*, *44*(3), 108–115. https://doi.org/10.1080/00219266.2010.9656206

Cicchetti, D. V. (1994a). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cicchetti, D. V. (1994b). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clouston, T. S. (1899). Stages of over-excitability, hypersensitiveness and mental explosiveness in children and their treatment by the bromides. *Scottish Med Surg J*, *4*, 481–490.

Coghill, D., Du, Y., Jiang, W., Xian, Lu, D., Qian, Y., Mulraney, M., & Su, L. (2021). A novel school-based approach to screening for attention deficit hyperactivity disorder. *European Child & Adolescent Psychiatry*, *31*(6), 909–917. https://doi.org/10.1007/s00787-021-01721-w

Cohen, L., Manion, L., & Morrision, K. (2011). *Research Methods in Education* (7th ed.). Routledge.

Cohen, L., Manion, L., & Morrison, K. R. B. (2018). *Research methods in education* (7th ed.). Routledge.

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1997). Revision and Restandardization of the Conners Teacher Rating Scale (CTRS-R): Factor Structure, Reliability, and Criterion Validity. *Journal of Abnormal Child Psychology*, *26*(4), 279–291.

Conrad, P., & Bergey, M. R. (2014). The impending globalization of ADHD: Notes on the expansion and growth of a medicalized disorder. *Social Science & Medicine*, *122*, 31–43. https://doi.org/10.1016/j.socscimed.2014.10.019

Cook, M., Young, A., Taylor, D., & Bedford, A. P. (1998). Personality correlates of alcohol consumption. *Personality and Individual Differences*, *24*(5), 641–647. https://doi.org/10.1016/s0191-8869(97)00214-6

Crichton, A. (1798). *An inquiry into the nature and origin of mental derangement: comprehending a concise system of the physiology and pathology of the human mind and a history of the passion and their affects*. T. Cadell & W. Davies.

Curtis, D. F., Pisecco, S., Hamilton, R. J., & Moore, D. W. (2006). Teacher perceptions of classroom interventions for children with ADHD: A cross-cultural comparison of teachers in the United States and New Zealand. *School Psychology Quarterly*, *21*(2), 171–196. https://doi.org/10.1521/scpq.2006.21.2.171

Danforth, J. S., & DuPaul, G. J. (1996). Interrater reliability of teacher rating scales for children with attention-deficit hyperactivity disorder. *Journal of Psychopathology and Behavioral Assessment*, *18*(3), 227–237. https://doi.org/10.1007/bf02229046

de Ramírez, R. D., & Shapiro, E. S. (2005). Effects of student ethnicity on judgments of ADHD symptoms among Hispanic and White teachers. *School Psychology Quarterly*, *20*(3), 268–287. https://doi.org/10.1521/scpq.2005.20.3.268

del Campo, N., Chamberlain, S. R., Sahakian, B. J., & Robbins, T. W. (2011). The Roles of Dopamine and Noradrenaline in the Pathophysiology and Treatment of Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *69*(12), e145–e157. https://doi.org/10.1016/j.biopsych.2011.02.036

Demaray, M. K., Elting, J., & Schaefer, K. (2003). Assessment of Attention-Deficit/Hyperactivity Disorder (ADHD): A comparative evaluation of five, commonly

used, published rating scales. *Psychology in the Schools*, *40*(4), 341–361. https://doi.org/10.1002/pits.10112

Department of Education, Class Size for Pupils in Years 1-4 (Foundation Stage and Key Stage 1) 2 (2011).

The Department of Education. (2001). Notice from the Ministry of Education on the Issuance of the 'Outline of Curriculum Reform for Compulsory Education (Trial)' - Official Website of the Ministry of Education of the People's Republic of China. Moe.gov.cn. http://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/200106/t20010608_167343.html

The Department of Education. (2020). Development of National Education in 2020 - Official Website of the Ministry of Education of the People's Republic of China. Moe.gov.cn. http://www.moe.gov.cn/jyb_sjzl/s5990/202111/t20211115_579974.html

Di, X., & McEwan, H. (2017). On teaching and learning. In *Chinese philosophy on teaching and learning: Xueji in the twenty-first century* (pp. 9–18). essay, State University of New York Press.

Ding, M., Li, Y., Li, X., & Kulm, G. (2008). Chinese teachers' perceptions of students' classroom misbehaviour. *Educational Psychology*, *28*(3), 305–324. https://doi.org/10.1080/01443410701537866

Dobria, L. (2011). *Longitudinal rater modeling with Splines* (dissertation). ProQuest LLC., Michigan, United States.

Doernberg, E., & Hollander, E. (2016). Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. *CNS Spectrums*, *21*(4), 295–299. https://doi.org/10.1017/s1092852916000262

Du, W., Yin, Q., Ma, C., & LiCheng, Z. (2003). A Comparison of Ratings on ADHD Symptoms by teachers in Guangzhou, HongKong and England. *Chinese Journal of Behavioural Medical Science*, *12*(5), 536–538.

Duckworth, A. L. (2011). The significance of self-control. *Proceedings of the National Academy of Sciences*, *108*(7), 2639–2640. https://doi.org/10.1073/pnas.1019725108

Duckworth, A. L., Kirby, T. A., Gollwitzer, A., & Oettingen, G. (2013). From Fantasy to Action: Mental Contrasting with Implementation Intentions (MCII) Improves Academic Performance in Children. *Social Psychological and Personality Science*, *4*(6), 745–753. https://doi.org/10.1177/1948550613476307

Duckworth, Angela L., Shulman, E. P., Mastronarde, A. J., Patrick, S. D., Zhang, J., & Druckman, J. (2015). Will not want: Self-control rather than motivation explains the female advantage in report card grades. *Learning and Individual Differences*, *39*, 13–23. https://doi.org/10.1016/j.lindif.2015.02.006

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*, *7*(4), 435–453. https://doi.org/10.1093/clipsy.7.4.435

Dyson, L. (1998). video of a boy and a girl diagnosed as having ADHD. *University of Victoria*.

Eckes, T. (2019). Many-facet Rasch Measurement : Implications for rater-mediated language assessment. In *Quantitative Data Analysis for Language Assessment Volume I: Fundamental Techniques* (1st ed., Vol. 7, p. 167). essay, Routledge.

Egger, H., Kondo, D., & Angold, A. (2006). The Epidemiology and Diagnostic Issues in Preschool Attention-Deficit/Hyperactivity Disorder. *Infants & Young Children*, *19*(2), 109–122. https://doi.org/10.1097/00001163-200604000-00004

Eisenberg, N., Fabes, R. A., Shepard, S. A., Murphy, B. C., Guthrie, I. K., Jones, S., Friedman, J., Poulin, R., & Maszk, P. (1997). Contemporaneous and Longitudinal Prediction of Children's Social Functioning from Regulation and Emotionality. *Child Development*, *68*(4), 642. https://doi.org/10.2307/1132116

Engelhard, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement*, *33*(1), 56–70. https://doi.org/10.1111/j.1745-3984.1996.tb00479.x

Ent, M. R., & Baumeister, R. F. (2014). Obedience, Self-Control, and the Voice of Culture. *Journal of Social Issues*, *70*(3), 574–586. https://doi.org/10.1111/josi.12079

Epstein, J. N., Willoughby, M., Valencia, E. Y., Tonev, S. T., Abikoff, H. B., Arnold, L. E., & Hinshaw, S. P. (2005). The Role of Children's Ethnicity in the Relationship Between Teacher Ratings of Attention-Deficit/Hyperactivity Disorder and Observed Classroom Behavior. *Journal of Consulting and Clinical Psychology*, *73*(3), 424–434. https://doi.org/10.1037/0022-006x.73.3.424

Erguvan, I. D., & Aksu Dunya, B. (2020). Analyzing rater severity in a freshman composition course using many Facet Rasch measurement. *Language Testing in Asia*, *10*(1). https://doi.org/10.1186/s40468-020-0098-3

Evertson, C. M., & Weinstein, C. S. (2006). *Handbook of classroom management* (1st ed.). Lawrence Erlbaum Associates.

Fabes, R. A., Eisenberg, N., Jones, S., Smith, M., Guthrie, I., Poulin, R., Shepard, S., & Friedman, J. (1999). Regulation, Emotionality, and Preschoolers' Socially Competent Peer Interactions. *Child Development*, *70*(2), 432–442. https://doi.org/10.1111/1467-8624.00031

Faraone, S. V., Biederman, J., & Mick, E. (2005). The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine*, *36*(2), 159–165. https://doi.org/10.1017/s003329170500471x

Faraone, S. V., Sergeant, J., Gillberg, C., Biederman, J., Biederman, J., Gillberg, C., Sergeant, J., & Faraone, S. V. (2003). The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry*, *2*(2), 104–113.

Farkas, G. (2003). Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. *Annu. Rev. Sociol.*, *29*(1), 541–562. https://doi.org/10.1146/annurev.soc.29.010202.100023

Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, *34*(1), 79. https://doi.org/10.37546/jaltjj34.1-3

Feng, Y. (2021). *Overview of Pediatrics Medical Industry in China* (p. 20). Nanjing: TF Research Institute.

First, M. B. (2012). A practical prototypic system for psychiatric diagnosis: The ICD-11 Clinical Descriptions and Diagnostic Guidelines. *World Psychiatry*, *11*(1), 24–25. https://doi.org/10.1016/j.wpsyc.2012.01.022

Flynn, T. M., & Flynn, T. M. (1985). Development of self-concept, delay of gratification and self-control and disadvantaged preschool children's achievement gain. *Early Child Development and Care*, *22*(1), 65–72. https://doi.org/10.1080/0300443850220105

Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: The Prevalence of DSM-IV Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(10), 1203–1211. https://doi.org/10.1097/00004583-200310000-00011

Franke, B., Michelini, G., Asherson, P., Banaschewski, T., Bilbow, A., Buitelaar, J. K., Cormand, B., Faraone, S. V., Ginsberg, Y., Haavik, J., Kuntsi, J., Larsson, H., Lesch, K.-P., Ramos-Quiroga, J. A., Réthelyi, J. M., Ribases, M., & Reif, A. (2018). Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *European Neuropsychopharmacology*, *28*(10), 1059–1088. https://doi.org/10.1016/j.euroneuro.2018.08.001

Funke, F., & Reips, U.-D. (2012). Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. *Field Methods*, *24*(3), 310–327. https://doi.org/10.1177/1525822x12444061

García Murillo, L., Ramos-Olazagasti, M. A., Mannuzza, S., Castellanos, F. X., & Klein, R. G. (2016). Childhood Attention-Deficit/Hyperactivity Disorder and Homelessness: A 33-Year Follow-Up Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(11), 931–936. https://doi.org/10.1016/j.jaac.2016.07.772

Gershon, J. (2002). A Meta-Analytic Review of Gender Differences in ADHD. *Journal of Attention Disorders*, *5*(3), 143–154. https://doi.org/10.1177/108705470200500302

Ghanizadeh, A. (2013). Agreement between Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, and the proposed DSM-V attention deficit hyperactivity disorder diagnostic criteria: an exploratory study. *Comprehensive Psychiatry*, *54*(1), 7–10. https://doi.org/10.1016/j.comppsych.2012.06.001

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

Gizer, I., Ficks, C., & Waldman, I. (2009). Candidate gene studies of ADHD: a meta-analytic review. *Human Genetics*, *126*(1), 51–90. https://doi.org/10.1007/s00439-009-0694-x

Godfrey, E., Fuermaier, A. B., Tucha, L., Butzbach, M., Weisbrod, M., Aschenbrenner, S., & Tucha, O. (2020). Public perceptions of adult ADHD: Indications of stigma? *Journal of Neural Transmission*. https://doi.org/10.1007/s00702-020-02279-8

Goldstein, S., Goldstein, M., Jones, C. B., Braswell, L., & Sheridan, S. M. (1998). *Managing attention deficit hyperactivity disorder in children* (1st ed.). John Wiley & Sons.

Gomez, R. (2007). Australian Parent and Teacher Ratings of the DSM-IV ADHD Symptoms. *Journal of Attention Disorders*, *11*(1), 17–27. https://doi.org/10.1177/1087054706295665

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *J Child Psychol & Psychiat*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Graham, P., & Rutter, M. (1968). Organic Brain Dysfunction and Child Psychiatric Disorder. *BMJ*, *3*(5620), 695–700. https://doi.org/10.1136/bmj.3.5620.695

Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, G. (2005). *Mental health of children and young people in Great Britain*. Palgrave MacMillan.

Grietens, H., Onghena, P., Prinzie, P., Gadeyne, E., Van Assche, V., Ghesquière, P., & Hellinckx, W. (2004). Comparison of Mothers', Fathers', and Teachers' Reports on Problem Behavior in 5- to 6-Year-Old Children. *Journal of Psychopathology and Behavioral Assessment*, *26*(2), 137–146. https://doi.org/10.1023/b:joba.0000013661.14995.59

Gu, M. (2006). An analysis of the impact of traditional Chinese culture on Chinese education. *Frontiers of Education in China*, *1*(2), 169–190. https://doi.org/10.1007/s11516-006-0001-8

Gu, Q., & Liu, X. (2004). The Integration of Two Assessment Tropisms of ADHD and The Establishment of Assessment Model. *Chinese Journal of Special Education*, *51*(9), 92–95.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hamed, A. M., Kauer, A. J., & Stevens, H. E. (2015). Why the Diagnosis of Attention Deficit Hyperactivity Disorder Matters. *Frontiers in Psychiatry*, *6*. https://doi.org/10.3389/fpsyt.2015.00168

Hart, E. L., Lahey, B. B., Loeber, R., Applegate, B., & Frick, P. J. (1995). Developmental change in attention-deficit hyperactivity disorder in boys: A four-year longitudinal study. *Journal of Abnormal Child Psychology*, *23*(6), 729–749. https://doi.org/10.1007/bf01447474

Hartman, C. A., Rhee, S. H., Willcutt, E. G., & Pennington, B. F. (2007). Modeling Rater Disagreement for ADHD: Are Parents or Teachers Biased? *Journal of Abnormal Child Psychology*, *35*(4), 536–542. https://doi.org/10.1007/s10802-007-9110-y

He, M., Su, Z., Peng, R., & Fan, Y. (2014). Prevalence survey analysis of preschooler obstructive sleep apnea syndrome and attention deficit hyperactivity disorder in Liwan district of Guangzhou City. *Chin Community Doctors*, *30*, 136–138.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Heckman, James J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*(1–2), 114–128. https://doi.org/10.1016/j.jpubeco.2009.11.001

Hervey, A., Epstein, J., & Curry, J. (2004). Neuropsychology of Adults With Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Neuropsychology*, *18*(3), 485–503. https://doi.org/10.1037/0894-4105.18.3.485.supp

Hinshaw, S. P., & Scheffler, R. M. (2014). *The ADHD explosion* (1st ed.). Oxford University Press.

Hire, A. J., Ashcroft, D. M., Springate, D. A., & Steinke, D. T. (2015). ADHD in the United Kingdom: Regional and socioeconomic variations in incidence rates amongst children and adolescents (2004-2013). *Journal of Attention Disorders*, *22*(2), 134–142. https://doi.org/10.1177/1087054715613441

Ho, W.-C. (2006). Popular culture in mainland Chinese education. *International Education Journal*, *7*(3), 348–363.

Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind*. McGraw-Hill.

Holden, S. E., Jenkins-Jones, S., Poole, C. D., Morgan, C., Coghill, D., & Currie, C. J. (2013a). The prevalence and incidence, resource use and financial costs of treating people with attention deficit/hyperactivity disorder (ADHD) in the United Kingdom (1998 to 2010). *Child and Adolescent Psychiatry and Mental Health*, *7*(1), 34. https://doi.org/10.1186/1753-2000-7-34

Hosterman, S. J., DuPaul, G. J., & Jitendra, A. K. (2008). Teacher ratings of ADHD symptoms in ethnic minority students: Bias or behavioral difference? *School Psychology Quarterly*, *23*(3), 418–435. https://doi.org/10.1037/a0012668

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, *5*(1), 64–86. https://doi.org/10.1037/1082-989x.5.1.64

Hoza, B., Gerdes, A. C., Hinshaw, S. P., Arnold, L. E., Pelham, W. E., Molina, B. S. G., Abikoff, H. B., Epstein, J. N., Greenhill, L. L., Hechtman, L., Odbert, C., Swanson, J. M., & Wigal, T. (2004). Self-Perceptions of Competence in Children With ADHD and Comparison Children. *Journal of Consulting and Clinical Psychology*, *72*(3), 382–391. https://doi.org/10.1037/0022-006x.72.3.382

Hui, E. K. P., Sun, R. C. F., Chow, S. S., & Chu, M. H. (2011). Explaining Chinese students' academic motivation: Filial piety and self-determination. *Educational Psychology*, *31*(3), 377–392. https://doi.org/10.1080/01443410.2011.559309

Hung, S.-P., Chen, P.-H., & Chen, H.-C. (2012). Improving creativity performance assessment: A rater effect examination with many facet Rasch model. *Creativity Research Journal*, *24*(4), 345–357. https://doi.org/10.1080/10400419.2012.730331

Hurtig, T., Ebeling, H., Taanila, A., Miettunen, J., Smalley, S. L., Mcgough, J. J., Loo, S. K., Jarvelin, M.-R., & Moilanen, I. K. (2007). ADHD Symptoms and Subtypes: Relationship Between Childhood and Adolescent Symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(12), 1605–1613. https://doi.org/10.1097/chi.0b013e318157517a

J. Anthony, C., Styck, K. M., Cooke, E., Martel, J. R., & E. Frye, K. (2021). Evaluating the impact of rater effects on behavior rating scale score validity and utility. *School Psychology Review*, *51*(1), 25–39. https://doi.org/10.1080/2372966x.2020.1827681

Jackson, D. A., & King, A. R. (2004). Gender Differences in the Effects of Oppositional Behavior on Teacher Ratings of ADHD Symptoms. *Journal of Abnormal Child Psychology*, *32*(2), 215–224. https://doi.org/10.1023/b:jacp.0000019772.71251.ff

James, W. (1890). *The principles of psychology*. Henry Holt and Company.

Jensen, A. R. (1980). *Bias in mental testing* (1st ed.). Free Press.

Ji, Z. (2010). Wushu Cartoon: The New Route International Communication for Wushu Culture. *Journal of Nanjing Institute of Physical Education (Social Science)*, *6*.

Jin, W., Du, Y., & Zhong, X. (2010). Survey on Awareness of Primary and Middle School Teachers about Attention Deficit Hyperactivity Disorder in Shanghai. *China Journal of Health Psychology*, *3*(24).

John, O. P., Robin, R. W., Robin, R. W., & John, O. P. (2021). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In *Handbook of personality: Theory and research* (4th ed., pp. 102–138). essay, The Guilford Press.

Johnson, L. A., & Safranek, S. (2005). What is the most effective treatment for ADHD in children? *Clinical Inquiries, (MU)*.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402. https://doi.org/10.4097/kjae.2013.64.5.402

Kaptan, F., & İzgi, Ü. (2014). The Effect of Use Concept Cartoons Attitudes of First Grade Elementary Students towards Science and Technology Course. *Procedia - Social and Behavioral Sciences*, *116*, 2307–2311. https://doi.org/10.1016/j.sbspro.2014.01.564

Kazda, L., Bell, K., Thomas, R., McGeechan, K., Sims, R., & Barratt, A. (2021). Overdiagnosis of attention-deficit/hyperactivity disorder in children and adolescents. *JAMA Network Open*, *4*(4). https://doi.org/10.1001/jamanetworkopen.2021.5335

Kemnitz, T. M. (1973). The Cartoon as a Historical Source. *Journal of Interdisciplinary History*, *4*(1), 81. https://doi.org/10.2307/202359

Kendall, P. C., & Wilcox, L. E. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology*, *47*(6), 1020–1029. https://doi.org/10.1037//0022-006x.47.6.1020

Kim, E., Akers, R. L., & Yun, M. (2013). A Cross-Cultural Test of Social Structure and Social Learning: Alcohol Use among South Korean Adolescents. *Deviant Behavior*, *34*(11), 895–915. https://doi.org/10.1080/01639625.2013.782787

Kitayama, S., & Cohen, D. (2020). *Handbook of Cultural Psychology*. The Guilford Press.

Kokkinos, C. M., Panayiotou, G., & Davazoglou, A. M. (2004). Correlates of teacher appraisals of student behaviors. *Psychology in the Schools*, *42*(1), 79–89. https://doi.org/10.1002/pits.20031

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lagrange, T. C., Silverman, R. A., Silverman, R. A., & Lagrange, T. C. (1999). Low self-control and opportunity: testing the general theory of crime as an explanation for gender differences in delinquency. *Criminology*, *37*(1), 41–72. https://doi.org/10.1111/j.1745-9125.1999.tb00479.x

Lahey, B. B., Pelham, W. E., Chronis, A., Massetti, G., Kipp, H., Ehrhardt, A., & Lee, S. S. (2006). Predictive validity of ICD-10 hyperkinetic disorder relative to DSM-IV attention-deficit/hyperactivity disorder among younger children. *Journal of Child*

Psychology and Psychiatry, *47*(5), 472–479. https://doi.org/10.1111/j.1469-7610.2005.01590.x

Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM-IV Subtypes of ADHD From Preschool Through Elementary School. *Archives of General Psychiatry*, *62*(8), 896. https://doi.org/10.1001/archpsyc.62.8.896

Lai, A., Cui, M., & Yin, Q. (2002). A Comparison of Assessment on Symptoms of ADHD by Chinese and British Teachers. *Chinese Mental Health Journal*, *2002*(12).

Lambert, M. C., Puig, M., Lyubansky, M., Rowan, G. T., & Winfrey, T. (2001). Adult Perspectives on Behavior and Emotional Problems in African American Children. *Journal of Black Psychology*, *27*(1), 64–85. https://doi.org/10.1177/0095798401027001004

Landers, R. (2015). Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS. *The Winnower*, *9*. https://doi.org/10.15200/winn.143518.81744

Larrson, J.-O., Larrson, H., & Lichtenstein, P. (2004). Genetic and Environmental Contributions to Stability and Change of ADHD Symptoms Between 8 and 13 Years of Age: A Longitudinal Twin Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(10), 1267–1275. https://doi.org/10.1097/01.chi.0000135622.05219.bf

Lebowitz, M. S. (2013). Stigmatization of ADHD: A Developmental Review. *Journal of Attention Disorders*, *20*(3), 199–205. https://doi.org/10.1177/1087054712475211

Lee, S. I., Schachar, R. J., Chen, S. X., Ornstein, T. J., Charach, A., Barr, C., & Ickowicz, A. (2008). Predictive validity of DSM-IV and ICD-10 criteria for ADHD and hyperkinetic

disorder. *Journal of Child Psychology and Psychiatry*, *49*(1), 70–78. https://doi.org/10.1111/j.1469-7610.2007.01784.x

Leff, S. S., Crick, N. R., Angelucci, J., Haye, K., Jawad, A. F., Grossman, M., & Power, T. J. (2006). Social Cognition in Context: Validating a Cartoon-Based Attributional Measure for Urban Girls. *Child Development*, *77*(5), 1351–1358. https://doi.org/10.1111/j.1467-8624.2006.00939.x

Leiner, M., Handal, G., & Williams, D. (2004). Patient communication: a multidisciplinary approach using animated cartoons. *Health Education Research*, *19*(5), 591–595. https://doi.org/10.1093/her/cyg079

Li, D., Sham, P., Owen, M., & He, L. (2006). Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). *Human Molecular Genetics*, *15*(14), 2276–2284. https://doi.org/10.1093/hmg/ddl152

Lian, T., Hui-jing, S., & Jiajie, Z. (2013). Prevalence of ADHD in children of China：a systematic review and meta analysis. *Chinese Journal of Public Health*, *29*(9), 1279–1283.

Linacre, J. M. (2021). *Missing data*. Winsteps.com. https://www.winsteps.com/winman/missingdata.htm#:~:text=Winsteps%20does%20not%20require%20complete%20data%20in%20order%20to%20make%20estimates.&text=The%20observed%20and%20expected%20marginal%20scores%20are%20obtained%20by%20summing,of%20parameter%20estimates%2C%20see%20RSA.

Liu, A., Xu, Y., Yan, Q., & Tong, L. (2018). The Prevalence of Attention Deficit/Hyperactivity Disorder among Chinese Children and Adolescents. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-29488-2

Liu, H., Liu, C., Chang, F., & Loyalka, P. (2016). Implementation of teacher training in China and its policy implications. *China & World Economy*, *24*(3), 86–104. https://doi.org/10.1111/cwe.12160

Liu, X., Wang, Z., & Chen, H. (2002). An investing of ADHD children using multiple assessments. *Chinese Journal of Special Education*, *34*(2), 65–69.

Livingston, R., & Livingston, R. (1999). Cultural Issues in Diagnosis and Treatment of ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(12), 1591–1594. https://doi.org/10.1097/00004583-199912000-00021

Livingstone, L. T., Coventry, W. L., Corley, R. P., Willcutt, E. G., Samuelsson, S., Olson, R. K., & Byrne, B. (2016). Does the Environment Have an Enduring Effect on ADHD? A Longitudinal Study of Monozygotic Twin Differences in Children. *Journal of Abnormal Child Psychology*, *44*(8), 1487–1501. https://doi.org/10.1007/s10802-016-0145-9

Lodhi, M. A., Ibrar, S., Shamim, M., & Naz, S. (2018). Linguistic Analysis of Selected TV Cartoons and Its Impact on Language Learning. *International Journal of English Linguistics*, *8*(5), 247. https://doi.org/10.5539/ijel.v8n5p247

Longshore, D. (1998). Self-Control and Criminal Opportunity: A Prospective Test of the General Theory of Crime. *Social Problems*, *45*(1), 102–113. https://doi.org/10.2307/3097145

Loughran, S. B. (2003). Agreement and Stability of Teacher Rating Scales for Assessing ADHD in Preschool. *Early Childhood Education Journal*, *30*(4), 247–253.

Luk, S. L., Leung, P. W. L., & Lee, P. L. M. (1988). CONNERS' TEACHER RATING SCALE IN CHINESE CHILDREN IN HONG KONG. *J Child Psychol & Psychiat*, *29*(2), 165–174. https://doi.org/10.1111/j.1469-7610.1988.tb00701.x

Madsen, K. B., Ersbøll, A. K., Olsen, J., Parner, E., & Obel, C. (2015). Geographic analysis of the variation in the incidence of ADHD in a country with free access to healthcare: A danish cohort study. *International Journal of Health Geographics*, *14*(1). https://doi.org/10.1186/s12942-015-0018-4

Maniadaki, K., & Kakouros, E. (2017). *The complete guide to ADHD: Natue, Diagnosis and Treatment* (1st ed.). Routledge.

Mannuzza, S., Klein, R. G., Bessler, A., Malloy, P., & LaPadula, M. (1998). Adult Psychiatric Status of Hyperactive Boys Grown Up. *American Journal of Psychiatry*, *155*(4), 493–498. https://doi.org/10.1176/ajp.155.4.493

Martel, M. M., Schimmack, U., Nikolas, M., & Nigg, J. T. (2015). Integration of symptom ratings from multiple informants in ADHD diagnosis: A psychometric model with clinical utility. *Psychological Assessment*, *27*(3), 1060–1071. https://doi.org/10.1037/pas0000088

Martinez-Badía, J. (2015). Who says this is a modern disorder? The early history of attention deficit hyperactivity disorder. *World Journal of Psychiatry*, *5*(4), 379. https://doi.org/10.5498/wjp.v5.i4.379

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/bf02296272

MATHER, B. (2012). The social construction and reframing of attention-deficit/hyperactivity disorder. *Ethical Human Psychology & Psychiatry*, *14*(1), 15–26.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, *101*(3), 689–704. https://doi.org/10.1037/a0014240

Maxwell, E., Simmons, M., Franklin, L., Arnold, J., & Pall, H. (2014). Impact of educational cartoon on pediatric bowel preparation quality at time of Colonoscopy. *Global Pediatric Health*, *1*. https://doi.org/10.1177/2333794x14548199

Merrell, C., Styles, I., Jones, P., Tymms, P., & Wildy, H. (2013). Cross-country comparisons of inattentive, hyperactive and impulsive behaviour in school-based samples of young children. *International Research in Early Childhood Education*, *4*(1), 1–17.

Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Mitsis, E. M. E. M., Mckay, K. K., Schulz , K. P., Newcorn , J. H., Halperin , J. M., Halperin , J. M., Newcorn , J. H., Schulz , K. P., Mckay, K. K., & Mitsis, E. M. E. M. (2000). Parent–Teacher Concordance for DSM-IV Attention-Deficit/Hyperactivity Disorder in a Clinic-Referred Sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(3), 308–313. https://doi.org/10.1097/00004583-200003000-00012

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Moldavsky, M., Groenewald, C., Owen, V., & Sayal, K. (2012). Teachers' recognition of children with ADHD: role of subtype and gender. *Child and Adolescent Mental Health*, *18*(1), 18–23. https://doi.org/10.1111/j.1475-3588.2012.00653.x

Molina, B. S. G., Pelham, W. E., Blumenthal, J., & Galiszewski, E. (1998). Agreement among teachers' behavior ratings of adolescents with a childhood history of attention deficit hyperactivity disorder. *Journal of Clinical Child Psychology*, *27*(3), 330–339. https://doi.org/10.1207/s15374424jccp2703_9

Moon, S. (2011). Cultural perspectives on attention deficit hyperactivity disorder: a comparison between Korea and the US. *Journal of International Business and Cultural Studies*, *6*(1).

Morris, G. D. (2006). Self-Control, Native Traditionalism, and Native American Substance Use: Testing the Cultural Invariance of a General Theory of Crime. *Crime & Delinquency*, *52*(4), 572–598. https://doi.org/10.1177/0011128705282988

MTA Cooperative Group. (1999). A 14-Month Randomized Clinical Trial of Treatment Strategies for Attention-Deficit/Hyperactivity Disorder. *Archives of General Psychiatry*, *56*(12), 1073. https://doi.org/10.1001/archpsyc.56.12.1073

Mueller, C. W., Mann, E. M., Thanapum, S., Humris, E., Ikeda, Y., Takahashi, A., Tai Tao, K., & Ling Li, B. (1995). Teacher's ratings of disruptive behavior in five countries. *Journal of Clinical Child Psychology*, *24*(4), 434–442. https://doi.org/10.1207/s15374424jccp2404_7

Mulholland, S. M., Cumming, T. M., & Jung, J. Y. (2015). Teacher attitudes towards students who exhibit ADHD-type behaviours. *Australasian Journal of Special Education*, *39*(1), 15–36. https://doi.org/10.1017/jse.2014.18

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Järvinen, L. (2012). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *British Journal of Educational Psychology*, *82*(2), 185–206. https://doi.org/10.1111/j.2044-8279.2010.02017.x

Murphy, B. C., & Eisenberg, N. (1997). Young Children's Emotionality, Regulation and Social Functioning and Their Responses When They Are Targets of a Peer's Anger. *Social Development*, *6*(1), 18–36. https://doi.org/10.1111/j.1467-9507.1997.tb00092.x

Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology*, *71*(1), 39–44. https://doi.org/10.1037/0021-9010.71.1.39

Murray, D. W., Kollins, S. H., Hardy, K. K., Abikoff, H. B., Swanson, J. M., Cunningham, C., Vitiello, B., Riddle, M. A., Davies, M., Greenhill, L. L., McCracken, J. T., McGough, J. J., Posner, K., Skrobala, A. M., Wigal, T., Wigal, S. B., Ghuman, J. K., & Chuang, S. Z. (2007). Parent versus Teacher Ratings of Attention-Deficit/Hyperactivity Disorder Symptoms in the Preschoolers with Attention-Deficit/Hyperactivity Disorder

Treatment Study (PATS). *Journal of Child and Adolescent Psychopharmacology*, *17*(5), 605–619. https://doi.org/10.1089/cap.2007.0060

Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, *5*(2), 189–227.

National curriculum in England: framework for key stages 1 to 4. GOV.UK. (2014). https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4/the-national-curriculum-in-england-framework-for-key-stages-1-to-4

Neumann, H. (1859). *Lehrbuch der Psychiatrie*. Enke Verlag.

NHS, N. (2018). *ADHD Diagnosis*. nhs.uk. https://www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/diagnosis/

Ni, X., Zhang, Y., Han, X., Lei, S., Sun, J., & Zhou, R. (2014). Traditional Chinese medicine in the treatment of ADHD: a review. *Child Adolesc Psychiatr Clin N Am*, *23*(4), 853–81.

NICE. (2018). *Attention deficit hyperactivity disorder: diagnosis and management | Guidance and guidelines | NICE*. Nice.org.uk. https://www.nice.org.uk/guidance/NG87

Norvilitis, J. M., & Fang, P. (2005). Perceptions of ADHD in China and the United States: A preliminary study. *Journal of Attention Disorders*, *9*(2), 413–424. https://doi.org/10.1177/1087054705281123

Ohan, J. L., Visser, T. A. W., Strain, M. C., & Allen, L. (2011). Teachers' and education students' perceptions of and reactions to children with and without the diagnostic label

"ADHD." *Journal of School Psychology*, *49*(1), 81–105. https://doi.org/10.1016/j.jsp.2010.10.001

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

O'Neill, S., Schneiderman, R. L., Rajendran, K., Marks, D. J., & Halperin, J. M. (2013). Reliable Ratings or Reading Tea Leaves: Can Parent, Teacher, and Clinician Behavioral Ratings of Preschoolers Predict ADHD at Age Six? *Journal of Abnormal Child Psychology*, *42*(4), 623–634. https://doi.org/10.1007/s10802-013-9802-4

O'Neill, T. A. (2017). An Overview of Interrater Agreement on Likert Scales for Researchers and Practitioners. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00777

Papageorgiou, V., Kalyva, E., Dafoulis, V., & Vostanis, P. (2008). Differences in Parents' and Teachers' Ratings of ADHD Symptoms and Other Mental Health Problems. *The European Journal of Psychiatry*, *22*(4). https://doi.org/10.4321/s0213-61632008000400003

Parens, E., & Johnston, J. (2009). Facts, values, and Attention-Deficit Hyperactivity Disorder (ADHD): an update on the controversies. *Child and Adolescent Psychiatry and Mental Health*, *3*(1), 1. https://doi.org/10.1186/1753-2000-3-1

Paulson, J. F., Buermeyer, C., & Nelson-Gray, R. O. (2005). Social rejection and ADHD in young adults: An analogue experiment. *Journal of Attention Disorders*, *8*(3), 127–135. https://doi.org/10.1177/1087054705277203

Pelham, W. E., Altszuler, A. R., Merrill, B. M., Raiker, J. S., Macphee, F. L., Ramos, M., Gnagy, E. M., Greiner, A. R., Coles, E. K., Connor, C. M., Lonigan, C. J., Burger, L.,

Morrow, A. S., Zhao, X., Swanson, J. M., Waxmonsky, J. G., & Pelham, W. E. (2022). The effect of stimulant medication on the learning of academic curricula in children with ADHD: A randomized crossover study. *Journal of Consulting and Clinical Psychology*, *90*(5), 367–380. https://doi.org/10.1037/ccp0000725

Peluso, T., Ricciardelli, L. A., & Williams, R. J. (1999). Brief report self-control in relation to problem drinking and symptoms of disordered eating. *Addictive Behaviors*, *24*(3), 439–442. https://doi.org/10.1016/s0306-4603(98)00056-2

Perold, M., Louw, C., & Kleynhans, S. (2010). Primary school teachers' knowledge and misperceptions of attention deficit hyperactivity disorder (ADHD). *South African Journal of Education*, *30*(3), 457–473. https://doi.org/10.15700/saje.v30n3a364

Perry, B., Wong, N.-Y., & Howard, P. (2006). Comparing primary and secondary mathematics teachers' beliefs about mathematics, mathematics learning and mathematics teaching in Hong Kong and Australia. *Mathematics Education in Different Cultural Traditions-A Comparative Study of East Asia and the West*, *9*, 435–448. https://doi.org/10.1007/0-387-29723-5_26

Pescosolido, B. A., Jensen, P. S., Martin, J. K., Perry, B. L., Olafsdottir, S., Fettes, D., Fettes, D., Olafsdottir, S., Perry, B. L., Martin, J. K., Jensen, P. S., & Pescosolido, B. A. (2008). Public Knowledge and Assessment of Child Mental Health Problems: Findings From the National Stigma Study-Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(3), 339–349. https://doi.org/10.1097/chi.0b013e318160e3a0

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-

regression analysis. *International Journal of Epidemiology*, *43*(2), 434–442. https://doi.org/10.1093/ije/dyt261

Polanczyk, G., & Rohde, L. A. (2007). Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. *Current Opinion in Psychiatry*, *20*(4), 386–392. https://doi.org/10.1097/yco.0b013e3281568d7a

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. *American Journal of Psychiatry*, *164*(6), 942–948. https://doi.org/10.1176/ajp.2007.164.6.942

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

President of People's Republic of China. (2006). *Compulsory Education Law of the People's Republic of China (Chairman Order No. 52)*. Gov.cn. http://www.gov.cn/flfg/2006-06/30/content_323302.htm

Primi, R., Silvia, P., Benedek, M., & Jauk, E. (2019). Applying many-facet rasch modeling in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 176–186. https://doi.org/10.31234/osf.io/kgzvj

R Raman, S., K C Man, K., Bahmanyar, S., Berard, A., Bilder, S., Boukhris, T., Bushnell, G., Crystal, S., & Furu, K. (2018). Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. *The Lancet Psychiatry*, *5*(10), 824–835.

Rahman, S., Robbins, T. W., Hodges, J. R., Mehta, M. A., Nestor, P. J., Clark, L., & Sahakian, B. J. (2006). Methylphenidate ('Ritalin') can Ameliorate Abnormal Risk-Taking Behavior in the Frontal Variant of Frontotemporal Dementia. *Neuropsychopharmacology*, *31*(3), 651–658. https://doi.org/10.1038/sj.npp.1300886

Rasch, Georg. (1993). *Probabilistic Models for Some Intelligence and Attainment Tests*. MESA Press.

Rasch, George. (1960). *Studies in Mathematical Psychology: Probabilistic Models for Some Intelligence and Attainment Tests* (1st ed.). Danmarks pædagogiske Institut.

Raudenbush, S., & Kasim, R. (1998). Cognitive Skill and Economic Inequality: Findings from the National Adult Literacy Survey. *Harvard Educational Review*, *68*(1), 33–80. https://doi.org/10.17763/haer.68.1.1j47150021346123

Renoux, C., Shin, J., Dell'Aniello, S., Fergusson, E., & Suissa, S. (2016). Prescribing trends of attention-deficit hyperactivity disorder (ADHD) medications in UK primary care, 1995–2015. *British Journal of Clinical Pharmacology*, *82*(3), 858–868. https://doi.org/10.1111/bcp.13000

Resnick, R. J. (2005). Attention deficit hyperactivity disorder in teens and adults: They don't all outgrow it. *Journal of Clinical Psychology*, *61*(5), 529–533. https://doi.org/10.1002/jclp.20117

Reynolds, W. M., & Stark, K. D. (1986). Self-control in children: A multimethod examination of treatment outcome measures. *Journal of Abnormal Child Psychology*, *14*(1), 13–23. https://doi.org/10.1007/bf00917218

Ridzuan, F., & Wan Zainon, W. M. N. (2019). A Review on Data Cleansing Methods for Big Data. *Procedia Computer Science*, *161*, 731–738. https://doi.org/10.1016/j.procs.2019.11.177

Roberts, B. W. (2009). Back to the future: Personality and Assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Rohde, L. A., Szobot, C., Polanczyk, G., Schmitz, M., Martins, S., & Tramontina, S. (2005). Attention-deficit/hyperactivity disorder in a diverse culture: Do research and clinical findings support the notion of a cultural construct for the disorder? *Biological Psychiatry*, *57*(11), 1436–1441. https://doi.org/10.1016/j.biopsych.2005.01.042

Rosenbaum, M. (1980). A schedule for assessing self-control behaviors: Preliminary findings. *Behavior Therapy*, *11*(1), 109–121. https://doi.org/10.1016/s0005-7894(80)80040-2

Rothbart, M. K., Rueda, M. R., Rueda, M. R., & Rothbart, M. K. (2005). The development of effortful control. In *Developing individuality in the human brain: A tribute to Michael* (1st ed., pp. 167–188). essay, American Psychological Association.

Rush, B. (1812). *Medical inquiries and observations upon the diseases of the mind*. Kimber and Richardson.

Russell, G., Rodgers, L. R., Ukoumunne, O. C., & Ford, T. (2014). Prevalence of Parent-Reported ASD and ADHD in the UK: Findings from the Millennium Cohort Study. *Journal of Autism and Developmental Disorders*, *44*(1), 31–40. https://doi.org/10.1007/s10803-013-1849-0

S Matza, L., Paramore, C., & Prasad, M. (2005). A review of the economic burden of ADHD. *Cost Effectiveness and Resource Allocation*, *3*(1), 5–14.

Sandra Kooij, J. J., Buitellaar, J. K., van den Orod, E. J., Furer, J. W., Th Rijnders, C. A., Hodiamont, P. P. G., Hodiamont, P. P. G., Th Rijnders, C. A., Furer, J. W., van den Orod, E. J., Buitellaar, J. K., & Sandra Kooij, J. J. (2005). Internal and external validity of Attention-Deficit Hyperactivity Disorder in a population-based sample of adults. *Psychological Medicine*, *35*(6), 817–827. https://doi.org/10.1017/s003329170400337x

Sanghera, R. S. (2016). Sir George Frederic Still (1868–1941): A 'father' to many children. *Journal of Medical Biography*, *24*(4), 474–477. https://doi.org/10.1177/0967772014537739

Saul, R. (2014). *ADHD Does not Exist: The Truth About Attention Deficit and Hyperactivity Disorder* (1st ed.). HarperCollins.

Sax, L., & Kautz, K. J. (2003). Who First Suggests the Diagnosis of Attention-Deficit/Hyperactivity Disorder? *The Annals of Family Medicine*, *1*(3), 171–174. https://doi.org/10.1370/afm.3

Sayal, K., & Taylor, E. (2005). Parent ratings of school behaviour in children at risk of attention deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica*, *111*(6), 460–465. https://doi.org/10.1111/j.1600-0447.2004.00487.x

Sayal, K., Goodman, R., & Ford, T. (2006). Barriers to the identification of children with attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *47*(7), 744–750.

Sayal, Kapil, & Goodman, R. (2009). Do parental reports of child hyperkinetic disorder symptoms at school predict teacher ratings? *European Child & Adolescent Psychiatry*, *18*(6), 336–344. https://doi.org/10.1007/s00787-009-0735-y

Sayal, Kapil, Hornsey, H., Warren, S., MacDiarmid, F., & Taylor, E. (2006). Identification of children at risk of Attention Deficit/Hyperactivity Disorder. *Social Psychiatry and Psychiatric Epidemiology*, *41*(10), 806–813. https://doi.org/10.1007/s00127-006-0100-0

Sayal, Kapil, Prasad, V., Daley, D., Ford, T., & Coghill, D. (2018). ADHD in children and young people: Prevalence, care pathways, and service provision. *The Lancet Psychiatry*, *5*(2), 175–186. https://doi.org/10.1016/s2215-0366(17)30167-0

Scheffler, R. M., Hinshaw, S. P., Modrek, S., & Levine, P. (2007). The Global Market For ADHD Medications. *Health Affairs*, *26*(2), 450–457. https://doi.org/10.1377/hlthaff.26.2.450

Schneider, H. (2006). Who Receives a Diagnosis of Attention-Deficit/ Hyperactivity Disorder in the United States Elementary School Population? *PEDIATRICS*, *117*(4), e601–e609. https://doi.org/10.1542/peds.2005-1308

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ane.0000000000002864

Schultz, B. K., & Evans, S. W. (2012). Sources of bias in teacher ratings of adolescents with ADHD. *Journal of Educational and Developmental Psychology*, *2*(1). https://doi.org/10.5539/jedp.v2n1p151

Sciutto, M. J., Nolfi, C. J., & Bluhm, C. (2004). Effects of Child Gender and Symptom Type on Referrals for ADHD by Elementary School Teachers. *Journal of Emotional and Behavioral Disorders*, *12*(4), 247–253. https://doi.org/10.1177/10634266040120040501

Scott, J. (2004). *How to write for animation*. Overlook.

Shen, J., Zhang, N., Zhang, C., Caldarella, P., Richardson, M. J., & Shatzer, R. H. (2009). Chinese elementary school teachers' perceptions of students' classroom behaviour problems. *Educational Psychology*, *29*(2), 187–201. https://doi.org/10.1080/01443410802654909

Sherman, J., Rasmussen, C., & Baydala, L. (2008). The impact of teacher factors on achievement and behavioural outcomes of children with Attention Deficit/Hyperactivity Disorder (ADHD): a review of the literature. *Educational Research*, *50*(4), 347–360. https://doi.org/10.1080/00131880802499803

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*(6), 978–986. https://doi.org/10.1037/0012-1649.26.6.978

Shurkin, J. (2015). Science and Culture: Cartoons to better communicate science. *Proceedings of the National Academy of Sciences*, *112*(38), 11741–11742. https://doi.org/10.1073/pnas.1515144112

Shweta, Bajpai, R. C., & Chaturvedi, H. K. (2015a). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology*, *41*(3), 20–27.

Shweta, Bajpai, R. C., & Chaturvedi, H. k. (2015b). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology*, *41*(3), 20–27.

Singh, I. (2008a). ADHD, culture and education. *Early Child Development and Care*, *178*(4), 347–361. https://doi.org/10.1080/03004430701321555

Singh, I. (2008b). Beyond polemics: science and ethics of ADHD. *Nature Reviews Neuroscience*, 78–78. https://doi.org/10.1038/nrn2537

Smith, M. (2017). Hyperactive Around the World? The History of ADHD in Global Perspective. *Social History of Medicine*, *30*(4), 767–787. https://doi.org/10.1093/shm/hkw127

Snider, V. E., Busch, T., & Arrowood, L. (2003). Teacher Knowledge of Stimulant Medication and ADHD. *Remedial and Special Education*, *24*(1), 46–56. https://doi.org/10.1177/074193250302400105

Sollie, H., Larsson, B., & Mørch, W.-T. (2013). Comparison of Mother, Father, and Teacher Reports of ADHD Core Symptoms in a Sample of Child Psychiatric Outpatients. *Journal of Attention Disorders*, *17*(8), 699–710. https://doi.org/10.1177/1087054711436010

Sonuga-Barke, E. J. S., Minocha, K., Taylor, E. A., & Sandberg, S. (1993). Inter-ethnic bias in teachers' ratings of childhood hyperactivity. *British Journal of Developmental Psychology*, *11*(2), 187–200. https://doi.org/10.1111/j.2044-835x.1993.tb00597.x

Sonuga-Barke, E. J., Schlotz, W., & Kreppner, J. (2010). Differentiating devel- opmental trajectories for conduct, emotion and peer problems following early deprivation. *Monographs of the Society for Research in Child Development*, *75*(1), 102–124. https://doi.org/10.1111/j.1540-5834.2010.00552.x

Stangor, C., Jhangiani, R., & Tarry, H. (2014). *Principles of social psychology* (1st ed.). BC Campus.

State Education Commission of the PRC. (1996). Primary School Management Regulations - Official Website of the Ministry of Education of the People's Republic of China. Moe.gov.cn.http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/199603/t19960309_81895.html

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: Evidence for a dual systems model. *Developmental Psychology*, *44*(6), 1764–1778. https://doi.org/10.1037/a0012955

Stevens, J., & Quittner, A. L. (1998). Factors influencing elementary school teachers' ratings of ADHD and ODD behaviors. *Journal of Clinical Child Psychology*, *27*(4), 406–414. https://doi.org/10.1207/s15374424jccp2704_4

Stewart, M. A. (1970). Hyperactive Children. *Scientific American*, *222*(4), 94–98. https://doi.org/10.1038/scientificamerican0470-94

Still, G., & Still, G. (1902). The Goulstonian lectures on some abnormal psychical conditions in Children. *The Lancet*, *1*, 1008–1013. https://doi.org/10.1016/s0140-6736(01)74984-7

Styck, K. M., Anthony, C. J., Flavin, A., Riddle, D., & LaBelle, B. (2021). Are ratings in the eye of the beholder? A non-technical primer on many Facet Rasch measurement to evaluate rater effects on teacher behavior rating scales. *Journal of School Psychology*, *86*, 198–221. https://doi.org/10.1016/j.jsp.2021.01.001

Swanson, J. M., Kinsbourne, M., Nigg, J., Lanphear, B., Stefanatos, G. A., Volkow, N., Taylor, E., Casey, B. J., Castellanos, F. X., & Wadhwa, P. D. (2007). Etiologic Subtypes of Attention-Deficit/Hyperactivity Disorder: Brain Imaging, Molecular Genetic and Environmental Factors and the Dopamine Hypothesis. *Neuropsychology Review*, *17*(1), 39–59. https://doi.org/10.1007/s11065-007-9019-9

Swensen, A. R., BIRNBAUM, H. G., SECNIK, K., MARYNCHENKO, M., GREENBERG, P., & CLAXTON, A. (2003). Attention-Deficit/Hyperactivity Disorder: Increased Costs for Patients and Their Families. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(12), 1415–1423. https://doi.org/10.1097/00004583-200312000-00008

Tan, C. (2017). Confucianism and education. *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.226

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *J Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Tannock, R. (2013). Rethinking ADHD and LD in DSM-5: Proposed Changes in Diagnostic Criteria. *Journal of Learning Disabilities*, *46*(1), 5–25. https://doi.org/10.1177/0022219412464341

Taylor, E. (2011). Antecedents of ADHD: a historical account of diagnostic concepts. *ADHD Attention Deficit and Hyperactivity Disorders*, *3*(2), 69–75. https://doi.org/10.1007/s12402-010-0051-x

Thapar, A., Cooper, M., Eyre, O., & Langley, K. (2013). Practitioner review: What have we learnt about the causes of ADHD? *Journal of Child Psychology and Psychiatry*, *54*(1), 3–16. https://doi.org/10.1111/j.1469-7610.2012.02611.x

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *PEDIATRICS*, *135*(4), e994–e1001. https://doi.org/10.1542/peds.2014-3482

Thome, J., Ehlis, A.-C., Fallgatter, A. J., Krauel, K., Lange, K. W., Riederer, P., Romanos, M., Taurines, R., Tucha, O., Uzbekov, M., & Gerlach, M. (2012). Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *The World Journal of Biological Psychiatry*, *13*(5), 379–400. https://doi.org/10.3109/15622975.2012.690535

Thorell, L. B., Chistiansen, H., Hammar, M., Berggren, S., Zander, E., & Bölte, S. (2018). Standardization and cross-cultural comparisons of the Swedish Conners 3® rating scales. *Nordic Journal of Psychiatry*, *72*(8), 613–620. https://doi.org/10.1080/08039488.2018.1513067

Timimi , S., Taylor, E., Taylor, E., & Timimi , S. (2003). ADHD is best understood as a cultural construct. *The British Journal of Psychiatry*, *184*(1), 8–9. https://doi.org/10.1192/bjp.184.1.8

Timimi, S., & Taylor, E. (2004). ADHD is best understood as a cultural construct. *British Journal of Psychiatry*, *184*(01), 8–9. https://doi.org/10.1192/bjp.184.1.8

Timimi, Sami. (2005). *Naughty boys: anti-social behaviour, ADHD and the role of culture* (1st ed.). Palgrave Macmillan.

Toledot, M. A., Yangco, R. T., & Espinosae, A. A. (2014). Media Cartoons: Effects on Issue Resolution in Environmental Education. *International Electronic Journal of Environmental Education*, *4*(1). https://doi.org/10.18497/iejee-green.99250

Torgersen, T., Gjervan, B., & Rasmussen, K. (2006). ADHD in adults: A study of clinical characteristics, impairment and comorbidity. *Nordic Journal of Psychiatry*, *60*(1), 38–43. https://doi.org/10.1080/08039480500520665

Treyvaud, K. (2014). Best practice when using the Strengths and Difficulties Questionnaire with extremely preterm children: are two informants better than one? *Dev Med Child Neurol*, *56*(5), 413–414. https://doi.org/10.1111/dmcn.12448

Tripp, G., Luk, S. L., Schaughency, E. A., & Singh, R. (1999). DSM-IV and ICD-10: A Comparison of the Correlates of ADHD and Hyperkinetic Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(2), 156–164. https://doi.org/10.1097/00004583-199902000-00014

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12067

Types of school. GOV.UK. (2022). https://www.gov.uk/types-of-school

Valera, E. M., Faraone, S. V., Murray, K. E., & Seidman, L. J. (2007). Meta-Analysis of Structural Imaging Findings in Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *61*(12), 1361–1369. https://doi.org/10.1016/j.biopsych.2006.06.011

Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., & Herbst, K. (2005). Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. *PLoS Medicine*, *2*(10), e267. https://doi.org/10.1371/journal.pmed.0020267

Vazsonyi, A. T., Clifford Wittekind, J. E., Belliston, L. M., & Van Loh, T. D. (2004). Extending the General Theory of Crime to "The East:" Low Self-Control in Japanese Late Adolescents. *Journal of Quantitative Criminology*, *20*(3), 189–216. https://doi.org/10.1023/b:joqc.0000037731.28786.e3

Wang, M., Luo, S., Jin, L., Wang, C., & Lin, J. (2002). The analysis of behavior problems in preschool children with attention deficit-hyperacitivity disorder. *Chin J Behav Med Sci*, *11*, 555–556.

Wang, T., Liu, K., Li, Z., Xu, Y., Liu, Y., Shi, W., & Chen, L. (2017). Prevalence of attention deficit/hyperactivity disorder among children and adolescents in China: a systematic review and meta-analysis. *BMC Psychiatry*, *17*(1). https://doi.org/10.1186/s12888-016-1187-9

Wehmeier, P. M., Schacht, A., & Barkley, R. A. (2010). Social and Emotional Impairment in Children and Adolescents with ADHD and the Impact on Quality of Life. *Journal of Adolescent Health*, *46*(3), 209–217. https://doi.org/10.1016/j.jadohealth.2009.09.009

Weiss, G., & Hechtman, L. T. (1993). *Hyperactive children grown up: ADHD in children, adolescents and adults*. Guilford Press.

West, C. (1848). *Lectures on the diseases of Infancy and Childhood ... Seventh edition ... enlarged*. London.

Wilcox, C. E., Washburn, R., & Patel, V. (2007). Seeking help for attention deficit hyperactivity disorder in developing countries: A study of parental explanatory models in Goa, India. *Social Science & Medicine*, *64*(8), 1600–1610. https://doi.org/10.1016/j.socscimed.2006.11.032

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biological Psychiatry*, *57*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006

Williams, J. (2008). Working toward a neurobiological account of ADHD: Commentary on Gail Tripp and Jeff Wickens, Dopamine transfer deficit. *Journal of Child Psychology and Psychiatry*, *49*(7), 705–711. https://doi.org/10.1111/j.1469-7610.2008.01921.x

Wills, T. A., DuHamel, K., & Vaccaro, D. (1995). Activity and mood temperament as predictors of adolescent substance use: Test of a self-regulation mediational model. *Journal of Personality and Social Psychology*, *68*(5), 901–916. https://doi.org/10.1037//0022-3514.68.5.901

Wong, Y. W., Kim, D., & Lee, J. (2012). Traditional Oriental Herbal Medicine for Children and Adolescents with ADHD: A Systematic Review. *Evidence-Based Complementary and Alternative Medicine*, *2012*, 1–15. https://doi.org/10.1155/2012/520198

Woo, S. M., & Keatinge, C. (2008). *Diagnosis and treatment of mental disorders across the lifespan* (1st ed.). John Wiley & Sons.

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach* (1st ed.). Educational Measurement Solutions.

Xiao, L. (2006). Bridging the Gap between Teaching Styles and Learning Styles: A Cross-Cultural Perspective. *Teaching English as a Second or Foreign Language*, *10*(3), 1–15.

Xu, X., Li, X.-M., Zhang, J., & Wang, W. (2017). Mental health-related stigma in China. *Issues in Mental Health Nursing*, *39*(2), 126–134. https://doi.org/10.1080/01612840.2017.1368749

Yang, L. (2018). Anger grows in china over school crowding. *The Economist*, *232*(13), 21–22.

Yu, J. H., Keown, C. F., & Jacobs, L. W. (1993). Attitude Scale Methodology: *Journal of International Consumer Marketing*, *6*(2), 45–64. https://doi.org/10.1300/j046v06n02_05

Yu, S., Kowitt, S. D., Fisher, E. B., & Li, G. (2018). Mental Health in China: Stigma, Family Obligations, and the Potential of Peer Support. *Community Mental Health Journal*, *54*(6), 757–764. https://doi.org/10.1007/s10597-017-0214-8

Yu-Cun, S., Yu-Feng, W., Xiao-Ling, Y., Xiao-Ling, Y., Yu-Feng, W., & Yu-Cun, S. (1985). An epidemiological investigation of minimal brain dysfunction in six elementary schools in Beijing. *Journal of Child Psychology and Psychiatry*, *26*(5), 777–787. https://doi.org/10.1111/j.1469-7610.1985.tb00591.x

Zhang, B., Zhu, Y., Deng, B., Wen, Q., Sun, Y., & Wang, X. (2011). Survey of ADHD among primary and middle school students in urban area of Guiyang city. *Matern Child Health Care China*, *26*, 4892–4894.

Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger/frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. *Journal of Educational Psychology*, *102*(1), 180–196. https://doi.org/10.1037/a0015908

Appendixes

Appendix 1

DSM-III Diagnostic Criteria for Attention Deficit Disorder (with or without Hyperactivity) ADD+H, ADD-H (APA, 1980)

The child displays, for his or her mental and chronological age, signs of developmentally inappropriate inattention, impulsivity, and hyperactivity. The signs must be reported by adults in the child's environment, such as parents and teachers. Because the symptoms are typically variable, they may not be observed directly by the clinician. When the reports of teachers and parents conflict, primary consideration should be given to the teacher reports because of greater familiarity with age-appropriate norms. Symptoms typically worsen in situations that require self-application, as in the classroom. Signs of the disorder may be absent when the child is in a new or a one-to-one situation.

The number of symptoms specified is for children between the ages of eight and ten, the peak age range for a referral. In younger children, more severe forms of the symptoms and a great number of symptoms are usually present. The opposite is true of older children.

A Inattention. At least three of the following:
- (1) Often fails to finish things he or she starts
- (2) Often doesn't seem to listen
- (3) Easily distracted
- (4) Has difficulty concentrating on schoolwork or other tasks requiring sustained attention
- (5) Has difficulty sticking to a play activity

B Impulsivity. At least three of the following:
- (1) Often acts before thinking
- (2) Shifts excessively from one activity to another
- (3) Has difficulty organizing working
- (4) Needs a lot of supervision
- (5) Frequently calls out in class
- (6) Has difficulty awaiting turn in games or group situation

C Hyperactivity. At least two of the following:

- (1) Runs about or climbs on things excessively
- (2) Has difficulty sitting still or fidgets excessively
- (3) Has difficulty staying seated
- (4) Moves about excessively during sleep
- (5) Is always "on the go" or acts as if "driven by a motor"

D Onset before the age of seven

E Duration of at least six months

F Not due to Schizophrenia, Affective Disorder, or Severe or Profound Mental Retardation.

314.00 Attention Deficit Disorder without Hyperactivity
All of the features are the same as those of Attention Deficit Disorder with Hyperactivity except for the absence of hyperactivity; the associated features and impairment are generally milder. Prevalence and familial pattern are unknown.

314.80 Attention Deficit Disorder, Residual Type
Diagnostic criteria for Attention Deficit Disorder, Residual Type
A. The individual once met the criteria for Attention Deficit Disorder with Hyperactivity. This information may come from the individual or from others, such as family members

B. Signs of hyperactivity are no longer present, but other signs of the illness have persisted to the present without periods of remission, as evidenced by signs of both attentional deficits and impulsivity (e.g., difficulty organizing work and completing tasks, difficulty concentrating, being easily distracted, making sudden decisions without thought of the consequences).

C. The symptoms of inattention and impulsivity result in some impairment in social or occupational functioning.

D. Not due to Schizophrenia, Affective Disorder, Severe or Profound Mental Retardation, or Schizotypal or Borderline Personality Disorders.

Appendix 2

DSM-III-R Diagnostic Criteria for Attention-Deficit Hyperactivity Disorder (APA, 1987)

A. A disturbance of at least six months during which at least eight of the following are present:
   (1) Often fidgets with hands or feet or squirms in seat (in adolescents, may be limited to the subjective feeling of restlessness)
   (2) Has difficulty remaining seated when required to do so
   (3) Is easily distracted by extraneous stimuli
   (4) Has difficulty awaiting turn in games or group situations
   (5) Often blurt out answers to questions before they have been completed
   (6) Has difficulty following through on instructions from others (not due to oppositional behaviour or failure of comprehension), e.g., fails to finish chores
   (7) Has difficulty sustained attention in tasks or play activities
   (8) Often shifts from one uncompleted activity to another
   (9) Has difficulty playing quietly
   (10) Often talks excessively
   (11) Often interrupts or intrudes on others, e.g., butts into other children's games
   (12) Often does not seem to listen to what is being said to him or her
   (13) Often loses things necessary for tasks or activities at school or at home (e.g., toy, pencils, books, assignments)
   (14) Often engages in physically dangerous activities without considering possible consequences (not for the purpose of thrillseeking), e.g., runs into street without looking

   Note: The above items are listed in descending order of discriminating power based on the data from a national field trial of the DSM-III-R criteria for Disruptive Behavior Disorders.

B. Onset before the age of seven

C. Does not meet the criteria for a Pervasive Developmental Disorder.

   Criteria for severity of Attention-Deficit Hyperactivity Disorder:
   Mild: Few if any, symptoms in excess of those required to make the diagnosis and only minimal or no impairment in school and social functioning.
   Moderate: Symptoms or functional impairment intermediate between " mild" and " severe".
   Severe: Many symptoms in excess of those required to make the diagnosis and pervasive impairment in functioning at home and school and with peers.

314.00 Undifferentiated Attention-deficit Disorder

This is a residual category for disturbances in which the predominant feature is the persistence of developmentally inappropriate and marked inattention that is not a symptom of another disorder, such as Mental Retardation or Attention-deficit Hyperactivity Disorder or of a disorganized and chaotic environment. Some of the disturbances that in DSM-III would have been categorised as Attention Deficit Disorder without Hyperactivity would be included in this category. Research is necessary to determine if this is a valid diagnostic category and, if so, how it should be defined.

Appendix 3

DSM-IV Diagnostic Criteria for Attention-Deficit Hyperactivity Disorder (APA, 1994)

A Either (1) or (2)

(1) Inattention: six (or more) of the following symptoms of inattention have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level:

(a) Often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities.
(b) Often has difficulty sustaining attention in tasks or play activities.
(c) Often does not seem to listen when spoken to directly
(d) Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (not due to oppositional behaviour or failure to understand instructions)
(e) Often has difficulty organizing tasks and activities.
(f) Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (such as schoolwork or homework)
(g) Often loses things necessary for tasks or activities (toy, school assignments, pencils, books, or tools)
(h) Is often easily distracted by extraneous stimuli
(i) Is often forgetful in daily activities

(2) Hyperactivity-impulsivity: six (or more) of the following symptoms of inattention have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level:

Hyperactivity

(a) Often fidgets with hands or feet or squirms in seat
(b) Often leaves seat in classroom or in other situations in which remaining seated is expected
(c) Often runs about or climbs excessively in situations in which it is inappropriate (in adolescents or adults, may be limited to subjective feelings of restlessness)
(d) Often has difficulty playing or engaging in leisure activities quietly
(e) If often "on the go" or often acts as if " driven by a motor"
(f) Often talks excessively

Impulsivity

(g) Often blurts out answers to questions before they have been completed
(h) Often has difficulty awaiting turn
(i) Often interrupts or intrudes on others (e.g., butts into conversations or games)

B Some hyperactive-impulsive or inattentive symptoms that caused impairment were present before age 7 years.

C	Some impairment from the symptoms is present in two or more settings (e.g., at school or work and at home).
D	There must be clear evidence of clinically significant impairment in social, academic, or occupational functioning.
E.	The symptoms do not occur exclusively during the course of a Pervasive Developmental Disorder, Schizophrenia, or other Psychotic Disorder and are not better accounted for by another mental disorder (e.g. Mood Disorder, Anxiety Disorder, Dissociated Disorder, or a Personality Disorder).
	314.01 Attention-Deficit/Hyperactivity Disorder, Combined Type: if both Criteria A1 and A2 are met for the past 6 months 314.00 Attention-Deficit/Hyperactivity Disorder, Predominantly Inattentive Type: if Criterion A1 is met but Criterion A2 is not met for the past 6 months 314.01 Attention-Deficit/Hyperactivity Disorder, Predominantly Hyperactive-Impulsive Type: if Criterion A2 is met but Criterion A1 is not met for the past 6 months Coding note: For individuals (especially adolescents and adults) who currently have symptoms that no longer meet full criteria, 'In Partial Remission' should be specified.

Appendix 4

DSM-V Diagnostic Criteria for Attention-Deficit Hyperactivity Disorder (APA, 2013)

A A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as characterized by (1) and/or (2):

Inattention: Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:

Note: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.

(a) Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
(b) Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).
(c) Often does not seem to listen when spoken to directly (e.g., the mind seems elsewhere, even in the absence of any obvious distraction).
(d) Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).
(e) Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).
(f) Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).
(g) Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).
(h) Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
(i) Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).

Hyperactivity and impulsivity: Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:

Note: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or a failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.

(a) Often fidgets with or taps hands or feet or squirms in seat.
(b) Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplaces, or in other situations that require remaining in place).
(c) Often runs about or climbs in situations where it is inappropriate. (Note: In adolescents or adults, may be limited to feeling restless.)
(d) Often unable to play or engage in leisure activities quietly.
(e) Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for an extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
(f) Often talks excessively.
(g) Often blurts out an answer before a question has been completed (e.g., completes people's sentences; cannot wait for turn in conversation).
(h) Often has difficulty waiting for his or her turn (e.g., while waiting in line).
(i) Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

B Several inattentive or hyperactive-impulsive symptoms were present prior to age 12 years.

C Several inattentive or hyperactive-impulsive symptoms are present in two or more settings (e.g., at home, school, or work; with friends or relatives; in other activities).

D There is clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning.

E The symptoms do not occur exclusively during schizophrenia or another psychotic disorder and are not better explained by another mental disorder (e.g., mood disorder, anxiety disorder, dissociative disorder, personality disorder, substance intoxication or withdrawal).

Specify whether:

314.01 (F90.2) Combined presentation: If both Criterion A1 (inattention) and Criterion A2 (hyperactivity-impulsivity) are met for the past 6 months.
314.00 (F90.0) Predominantly inattentive presentation: If Criterion A1 (inattention) is met but Criterion A2 (hyperactivity-impulsivity) is not met for the past 6 months.
314.01 (F90.1) Predominantly hyperactive/impulsive presentation: If Criterion A2 (hyperactivity impulsivity) is met, but Criterion A1 (inattention) is not met over the past 6 months.

Specify if:

In partial remission: When full criteria were previously met, fewer than the full criteria have been met for the past 6 months, and the symptoms still result in impairment in social, academic, or occupational functioning.

Specify current severity:

Mild: Few, if any, symptoms in excess of those required to make the diagnosis are present, and symptoms result in only minor functional impairments.
Moderate: Symptoms or functional impairment between "mild" and "severe" are present.
Severe: Many symptoms in excess of those required to make the diagnosis, or several symptoms that are particularly severe, are present, or the symptoms result in marked impairment in social or occupational functioning.

Appendix 5

Conners Rating Scales-Revised (CRS-Rs) (Conners, 1997)

Behaviours		
1. Inattentive, easily distracted	11. Leaves seat in classroom or in other situations in which remaining seated is expected	21. Runs about or climbs excessively in situations where it is inappropriate
2. Defiant	12. Fidgets with hands or feet or squirms in seat	22. Poor in arithmetic
3. Restless in the "squirmy" sense	13. Not reading up to par	23. Interrupts or intrudes on others (eg., butts into others' conversations or games)
4. Forgets things he/she has already learned	14. Short attention span	24. Has difficulty playing or engaging in leisure activities quietly
5. Disturbs other children	15. Argues with adults	25. Fails to finish things he/she starts
6. Actively defies or refuses to comply with adults' requests	16. Only pays attention to things he/she is really interested in	26. Does not follow through on instructions and fails to finish schoolwork (not due to oppositional behaviour or failure to understand instructions)
7. Is always "on the go" or acts as if driven by a motor	17. Has difficulty waiting for his/her turn	27. Excitable, impulsive
8. Poor in spelling	18. Lacks interest in schoolwork	28. Restless, always up and on the go
9. Cannot remain still	19. Distractibility or attention span a problem	
10. Spiteful or vindictive	20. Temper outbursts; explosive, unpredictable behaviour	

Appendix 6

Strengths and Difficulties Questionnaires (SDQ)

Behaviours		
1. Considerate of other people's feelings	10. Constantly fidgeting or squirming	19. Picked on or bullied by other children
2. Restless, overactive, cannot stay still for long	11. Has at least one good friend	20. Often volunteers to help others (parents, teachers, other children)
3. Often complains of headaches, stomach-aches or sickness	12. Often fights with other children or bullies them	21. Thinks things out before acting
4. Shares readily with other children (treats, toys, pencils etc.)	13. Often unhappy, down-hearted or tearful	22. Steals from home, school or elsewhere
5. Often has temper tantrums or hot tempers	14. Generally liked by other children	23. Gets on better with adults than with other children
6. Rather solitary, tends to play alone	15. Easily distracted, concentration wanders	24. Many fears, easily scared
7. Generally obedient, usually does what adults request	16. Nervous or clingy in new situations, easily loses confidence	25. Sees tasks through to the end, good attention span
8. Many worries, often seems worried	17. Kind to younger children	
9. Helpful if someone is hurt, upset or feeling ill	18. Often lies or cheats	

Appendix 7

ADHD Rating Scale-IV

Behaviours		
1. Fidgets with hands or feet or squirms in seat	7. Does not follow through on instructions. Eg. fails to finish work.	13. Loses things necessary for tasks or activities.
2. Fails to give close attention to details or makes careless mistakes in schoolwork.	8. Has difficulty playing quietly or engaging in quiet-time activities.	14. Blurts out answers before questions have been completed.
3. Has difficulty sustaining attention in tasks or play activities.	9. Has difficulty organizing tasks and activities.	15. Is easily distracted
4. Leaves seat in classroom or situations when expected to remain seated	10. Is “on the go” or acts as if “driven by a motor.”	16. Has difficulty waiting for own turn.
5. Does not seem to listen when spoken to directly.	11. Avoids tasks (e.g., schoolwork, homework) that require continued mental effort.	17. Is forgetful in daily activities.
6. Runs or climbs excessively in situations where it is inappropriate.	12. Talks excessively.	18. Interrupts or intrudes on others.

Appendix 8

Attention Deficit Disorder Evaluation Scales

Behaviours	
1.Rushes through assignments with little or no regard for accuracy or quality of work	31. Begins assignments before receiving directions or instructions or does not follow directions or instructions
2. Is easily distracted by other activities in the classroom, other students, the teacher, etc.	32. Does not wait for his/her turn in activities or games
3. Does not listen to what other students are saying	33. Grabs things away from others
4. Does not hear all of what is said (e.g., misses word endings, misses keywords such as "do not," etc.)	34. Blurts out answers without being called on
5. Does not direct attention or fails to maintain attention to important sounds in the immediate environment (e.g., teacher directions, public address system, etc.)	35. Interrupts the teacher (e.g., begins talking while the teacher is talking, goes up to the teacher while the teacher is working with other students, etc.)
6. Is unsuccessful in activities requiring listening (e.g., games, following oral directions, etc.)	36. Interrupts other students (e.g., talks while they are talking, makes noises, laughs, etc
7. Needs oral questions and directions frequently repeated (e.g., student says, "I don't understand," needs constant reminders, etc.)	37. Talks to others during quiet activity periods
8. Attends more successfully when close to the source of sound (e.g., when seated close to the teacher, etc.)	38. Moves about while seated, fidgets, squirms, etc.
9. Requires eye contact in order to listen successfully (e.g., one-to-one situation, etc.)	39. Appears restless (e.g., shifts position in seat, paces about, etc.)

10. Fails to demonstrate short-term memory skills (e.g., does not remember two- or three-step directions, does not remember materials needed for a task, etc.)	40. Is easily angered, annoyed, or upset (e.g., hits, argues, yells, throws things, cries, withdraws, etc.)
11. Fails to remember sequences (e.g., events in a daily routine, days of the week, months of the year, etc.)	41. Bothers other students who are trying to work, listen, etc.
12.Has difficulty concentrating (e.g., staying on an assigned task, following a conversation, etc.)	42 Makes unnecessary comments or noises in the classroom (e.g., talks to other students without permission, interrupts, makes fun of others, hums, taps, burps, etc.)
13.Loses place when reading (e.g., leaves out words, lines, or sentences when reading, etc.)	43.Makes unnecessary physical contact with others (e.g., touches, hugs, etc.)
14.Omits, adds, substitutes, or reverses letters, words, or sounds when reading	44. Is impulsive (e.g., reacts immediately to situations without thinking, is impatient, fails to wait for a turn or for assistance from an instructor, etc.)
15. Fails to copy letters, words, sentences, and numbers from a textbook, chalkboard, etc.	45. Fails to comply with teachers or other school personnel (e.g., refuses to do what he/she is told, goes on doing what he/she was doing, does the opposite of what he/she is told, etc.)
16. Omits, adds, or substitutes words when writing	46. Ignores consequences of his/her behavior (i.e., knows that his/her behavior will result in a negative consequence but engages in the behavior anyway)
17. Fails to complete homework assignments and return them to school	47. Fails to follow a routine (e.g., does things out of order, does not wait for an activity at the scheduled time, etc.)
18. Does not perform or complete classroom assignments during class time (e.g., does not perform the assignment or does not use the time provided, will go on to another assignment before completing the first, etc.)	48. Does not follow the rules of games

19. Is disorganized to the point of not having necessary materials, losing materials, failing to completed assignments, failing to follow the steps of the assignment in order, etc.	49. Leaves seat without permission
20. Completes assignments with little or no regard for neatness (e.g., rushes through tasks, does not care to do well, etc.)	50. Does not work in a group situation (e.g., does not stay on-task and work productively at a table with peers, at a desk with peers nearby, etc.)
21. Fails to perform assignments independently (e.g., continually asks for assistance or reassurance; does not begin, work on, or complete assignments without assistance; etc.)	51. Hops, skips and jumps when moving from one place to another instead of walking
22. Does not prepare for school assignments (e.g., does not study for tests or quizzes, does not read the assigned material, etc.)	52. Handles objects (e.g., twirls pencils, plays with things in desk, spins rulers on pencils, clicks ballpoint pens, repeatedly sharpens pencils, etc.)
23. Does not remain on-task (e.g., is more interested in other activities, sits and does nothing, etc.)	53. Talks beyond what is expected or at inappropriate times
24. Does not perform academically at his/her ability level (i.e., performs below ability level or at a failing level)	54. Does not wait appropriately for assistance from the instructor
25. Does not listen to or follow verbal directions	55. Does not adjust behavior to the expectations of different situations (e.g., gets excited at recess and does not settle down, etc.)
26. Fails to make appropriate use of study time (e.g., does not read, study, work on assignments, etc.)	56. Engages in inappropriate behaviours while seated (e.g., tips chair or desk, puts feet on the desk, touches others as they walk by, taps and makes noises, etc.)
27. Fails to follow necessary steps in math problems (e.g., does step in the wrong order, omits a step, etc.)	57. Becomes overexcited (e.g., loses control in group activities, becomes loud, etc.)

28. Does not read or follow written directions (e.g., instructions for homework, assignment directions, etc.)	58. Demonstrates inappropriate behavior when moving with a group (e.g., fails to stay in line, runs, pushes, etc.)
29. Changes from one activity to another without finishing the first, without putting things away, before it is time to move on, etc.	59. Moves about unnecessarily (e.g., leaves the seat, walks around the classroom, rocks, shakes the head, etc.)
30. Does not follow school rules (e.g., runs in hallways, throws food in the cafeteria, is disruptive in the library, etc.)	60. Engages in nervous habits (e.g., bites fingernails, twirls hair, chews inside of the cheek, chews pencils or pens, spins or twirls objects, etc.)

Appendix 9

Vanderbilt ADHD Diagnostic Teacher Rating Scale

Behaviours		
1. Fails to give attention to details or makes careless mistakes in schoolwork	13. Has difficulty playing or engaging in leisure activities quietly	25. Lies to obtain goods for favours or to avoid obligations (i.e., "cons" others)
2. Has difficulty sustaining attention to tasks or activities	14. Is "on the go" or often acts as if "driven by a motor"	26. Is physically cruel to people
3. Does not seem to listen when spoken to directly	15. Talks excessively	27. Has stolen items of nontrivial value
4. Does not follow through on instructions and fails to finish schoolwork (not due to oppositional behavior or failure to understand)	16. Blurts out answers before questions have been completed	28. Deliberately destroys others' property
5. Has difficulty organizing tasks and activities	17. Has difficulty waiting inline	29. Is fearful, anxious, or worried
6. Avoids, dislikes or is reluctant to engage in tasks that require 0 1 2 3 sustaining mental effort	18. Interrupts or intrudes on others (e.g., butts into conversations or games)	30. Is self-conscious or easily embarrassed
7. Loses things necessary for tasks or activities (school assignments, pencils, or books)	19. Loses temper	31. Is afraid to try new things for fear of making mistakes
8. Is easily distracted by extraneous stimuli	20. Actively defies or refuses to comply with adults' requests or rules	32. Feels worthless or inferior
9. Is forgetful in daily activities	21. Is angry or resentful	33. Blames self for problems, feels guilty
10. Fidgets with hands or feet or squirms in seat	22. Is spiteful and vindictive	34. Feels lonely, unwanted, or unloved; complains that "no one loves him/her"

11. Leaves seat in classroom or in other situations in which remaining seated is expected	23. Bullies, threatens, or intimidates others	35. Is sad, unhappy, or depressed
12. Runs about or climbs excessively in situations in which remaining seated is expected	24. Initiates physical fights	

Appendix 10

Copeland Symptom Checklist for Attention Deficit Disorders

Inattention/Distractibility	
1. A short attention span, especially for low-interest activities	5. Nicknames such as “Pacey” or “Dreamer”
2. Difficulty completing tasks	6. Engages in much activity, but accomplishes little
3. Daydreaming	7. Enthusiastic beginnings, but poor endings
4. Easily distracted	
Impulsivity	
1. Excitability	6. Excessively shifts from one activity to another
2. Low frustration tolerance	7.Difficulty in group situations which require patience and taking turns
3. Acts before thinking	8. Requires much supervision
4. Disorganization	9. Constantly in trouble for deeds of omission as well as deeds of commission
5. Poor planning ability	10. Frequently interrupts conversations: talks out turn
Activity level problems - Overactivity/Hyperactivity	
1. Restlessness - either fidgetiness or being constantly on the go	4. Excessive running, jumping and climbing
2. Diminished need for help	5. Motor restlessness during sleep, kicks covers off, moves around
3. Excessive talking	6. Difficulty staying seated at meals, in class, etc. Walks around class
Underactivity	
1. Lethargy	4.Inattention
2. Daydreaming, spaciness	5. Poor leadership

3. Failure to complete tasks	6. Difficulty in learning and performing
Non-Compliance	
1. Frequently disobeys	4. “Forgets” unintentionally
2. Argumentative	5. Uses “forgetting” as an excuse (intentional)
3. Disregards socially accepted standards of behavior	
Attention Getting Behavior	
1. Frequently needs to be the centre of attention	4. Behaves as the “Class Clown”
2. Constantly asks questions or interrupts	5. Uses bad or rude language to attract attention
3. Irritates and annoys siblings, peers, and adults	6. Engages in other negative behaviours to attract attention
Immaturity	
1. Behavior resembles that a younger child, Responses are typical of children 6 months to 2-plus years younger	3. Prefers younger children and relates better to them
2. Physical development is delayed	4. Emotional reactions are often immature
Poor achievement/Cognitive & Visual- Motor problems	
1. Underachieve relative to the ability	6. Academic work completed too quickly
2. Loses books, assignments, etc	7. Academic work completed too slowly
3. Auditory memory and auditory processing problems	8. “Messy” or “ Sloppy” written work; poor handwriting
4. Learning disabilities/learning problems	9. Poor memory for directions, instructions and role learning
5. Incomplete assignments	
Emotional difficulties	
1. Frequent and unpredictable mood swings	5. Low frustration tolerance
2. High levels of irritability	6. Temper tantrums, angry outbursts

3. Under reactive to pain/insensitive to danger	7. Moodiness
4. Easily overstimulated. Hard to calm down once over-excited	8. Low self-esteem
Poor peer relations	
1. Hits, bites or kicks other children	4. Avoids group activities; a loner
2. Difficulty following the rules of games and social interactions	5. Teases peers and siblings excessively
3. Rejected or avoided by peers.	6.Bullies or bosses other children

Appendix 11

Swanson, Nolan, and Pelham-IV Questionnaire

Behaviours	
1. Often fails to give close attention to details or makes careless mistakes in schoolwork and tasks	14. Often has difficulty playing or engaging in leisure activities quietly
2. Often has difficulty sustaining attention in tasks or play activities	15. Often is "on the go" or often acts as if "driven by a motor"
3. Often does not seem to listen when spoken to directly	16. Often talks excessively
4. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties	17. Often blurts out answers before questions have been completed
5. Often has difficulty organizing tasks and activities	18. Often has difficulty awaiting turn
6. Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework)	19 Often interrupts or intrudes on others (e.g., butts into conversations/games
7. Often loses things necessary for tasks or activities (e.g., toys, school assignments, pencils, books, or tools)	20. Often has difficulty sitting still, being quiet, or inhibiting impulses in the classroom or at home
8. Often is distracted by extraneous stimuli	21. Often loses temper
9. Often is forgetful in daily activities	22. Often argues with adults
10. Often has difficulty maintaining alertness, orienting to requests, or executing directions	23. Often actively defies or refuses adult requests or rules
11. Often fidgets with hands or feet or squirms in seat	24. Often deliberately does things that annoy other people
12. Often leaves seat in classroom or in other situations in which remaining seated is expected	25. Often blames others for his or her mistakes or misbehavior
13. Often runs about or climbs excessively in situations in which it is inappropriate	26. Often touchy or easily annoyed by others

Appendix 12

Intra-class correlation coefficient (ICC) of 40 groups of Chinese teachers

Raters Group	Can't wait	Fidgety	Inattentive	Dis-organised	Dis-tracted	Dis-ruptive	Group Average
t001, t002	.37	.15	.41	.11	-.08	.00	.16
t003-t007	.47	.56	.54	.34	.35	.40	.44
t008, t009	.32	.58	.59	.33	.44	.55	.47
t010, t011	-.46	.26	.70	.24	.39	-.11	.17
t012, t013	.62	.77	.58	.48	.73	.55	.62
t014, t015	.13	.33	.46	.00	.39	.70	.34
t016, t017	.48	-.03	.52	.39	.69	.39	.41
t019, t020	.39	.02	-.41	.39	.18	.00	.10
t022, t023	.96	.89	.71	.70	.75	.87	.81
t024, t025	.89	.85	.79	.82	.79	.75	.82
t026, t027	.94	.93	.76	.69	.80	.41	.76
t028, t029	.40	.66	.44	.83	.56	.28	.53
t031-t033	.78	.72	.81	.46	.76	.55	.68
t034, t035	-.43	-.41	.65	.26	.64	.66	.23
t036, t037	.85	.69	.90	.00	.75	.66	.64
t038, t039	.47	.73	.56	.23	.53	.54	.51
t040, t041	.71	.82	.78	.79	.71	.59	.73
t042, t043	.94	.94	.85	.89	.93	.79	.89
t049, t050	.62	.61	.80	.37	.94	.57	.65
t052, t053	.00	-.30	.00	.00	-.09	-.33	-.12
t054, t055	.83	.74	.64	.63	.48	.71	.67
t056-t058	.19	.51	.76	.70	.73	.63	.59
t060, t061	.88	.68	.89	.39	.69	.92	.74

t062-t065	.32	.53	.58	.26	.27	.49	.41
t066-t068	.44	.79	.47	.62	.61	.77	.62
t069, t070	.77	.75	.57	.24	.46	.64	.57
t071, t072	-.92	.87	.92	.86	.86	.94	.59
t073, t074	.38	.62	.88	.47	.27	.05	.45
t075, t076	.23	-.13	.61	.06	.46	.84	.35
t078, t079	.46	.27	.63	.21	-.09	.18	.28
t080, t081	.39	.73	.86	.86	.81	.73	.73
t082, t083	.70	.90	.89	.58	.91	.80	.80
t086, t087	.75	.53	.70	.61	.39	.54	.59
t088, t089	.90	.89	.87	.69	.92	.86	.86
t090, t091	.66	.80	.10	.84	.28	.84	.59
t093, t094	.70	.54	.82	.00	.74	.69	.58
t095-t097	.80	.89	.77	.82	.71	.80	.80
t100, t101	.56	.30	.27	.45	.10	.45	.36
t102, t103	.53	.69	.87	.58	.59	.38	.61
t104,4105	.49	.07	.87	.68	.69	.48	.55
Symptom Average	.49	.54	.64	.47	.55	.54	

Appendix 13

Consent form

Shaped by the past, creating the future

Durham University
School of Education

APPENDIX B

[10/02/16]

Participant Information Sheet

Title: The self-control behaviors of young children across English and Chinese cultures

You are invited to take part in a research study of the self-control behaviours of young children across English and Chinese cultures. Please read this form carefully and ask any questions you may have before agreeing to be in the study. The study is conducted by Lin Lu as part of her postgraduate studies at Durham University. This research project is supervised by Prof Peter Tymms and Dr. Christine Merrell (Peter.Tymms@cem.dur.ac.ukChristine.Merrell@cem.dur.ac.uk) from the School of Education at Durham University.

The purpose of this study is to find whether the expectations or standards of self-control behaviours of young children are different across English and Chinese cultures.If you agree to be in this study, you will be asked to watch some cartoons which relate to young children's behaviours and rate the behaviours represented in these cartoons. Meanwhile, you will be asked to rate 10 pupils of their self-control behaviours in your own class. Your participation in this study will take approximately 20 minutes.You are free to decide whether to participate. If you decide to participate, you are free to withdraw at any time without any negative consequences for you.All responses you give or other data collected will be kept confidential. The records of this study will be kept secure and private. All files containing any information you give are password protected. In any research report that may be published, no information will be included that will make it possible to identify you individually. There will be

no way to connect your name to your responses at any time during or after the study. If you have any questions, requests or concerns regarding this research, please contact me via email Lin Lu (lin.lu@durham.ac.uk) or by telephone at 07450909581

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 08/03/16)

Lin Lu

Leazes Road

Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311

www.durham.ac.uk

APPENDIX C

Declaration of Informed Consent

I agree to participate in this study, the purpose of which is to find the expectations or standards of self-control behaviours of young children in schools across English and Chinese cultures.

I have read the participant information sheet and understand the information provided.

I have been informed that I may decline to answer any questions or withdraw from the study without penalty of any kind.

I have been informed that all my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.

I have been informed that the investigator will answer any questions regarding the study and its procedures. Lin Lu School of Education, Durham University can be contacted via email: lin.lu@durham.ac.uk or telephone: 07450909581

I will be provided with a copy of this form for my records.

Any concerns about this study should be addressed to the School of Education Ethics Sub-Committee, Durham University via email to ed.ethics@durham.ac.uk.

Date Participant Name (please print) Participant Signature

I certify that I have presented the above information to the participant and secured his or her consent.

Date Signature of Investigator

Leazes Road

Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311

www.durham.ac.uk

Durham University is the trading name of the University of Durham

Appendix14

Teacher's Rating Scale of Cartoons

Teacher's Rating Scale of Cartoons

Background information of teacher Please tick the ☐ you choose

School of teaching _____

Class Teacher: ☐ Teaching Assistant: ☐ Gender: Female ☐ Male ☐

Years of teaching: _____ Number of Children in Y2 Class: _____

Example of rating

How do you rate the behaviour of Child A ?

Cartoon 1 Can't wait

What do you think of the behaviour of Child A ? 

Unacceptable

Better than expected

What do you think of the behaviour of Child B ? 

Unacceptable

Better than expected

What do you think of the behaviour of Child C ? 

Unacceptable

Better than expected

What do you think of the behaviour of Child D ? 

Unacceptable

Better than expected

What do you think of the behaviour of Child E ? 

Unacceptable

Better than expected

Cartoon 2 Fidgety

What do you think of the behaviour of Child A ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child B ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child C ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child D ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child E ? 

Unacceptable — Better than expected

Cartoon 3 Inattentive

What do you think of the behaviour of Child A ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child B ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child C ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child D ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child E ? 

Unacceptable — Better than expected

Cartoon 4 Disorganized

What do you think of the behaviour of Child A ? 


Unacceptable


Better than expected

What do you think of the behaviour of Child B ? 


Unacceptable


Better than expected

What do you think of the behaviour of Child C ? 


Unacceptable


Better than expected

What do you think of the behaviour of Child D ? 


Unacceptable


Better than expected

What do you think of the behaviour of Child E ? 


Unacceptable


Better than expected

Cartoon 5 Distracted

What do you think of the behaviour of Child A ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child B ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child C ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child D ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child E ? 

Unacceptable — Better than expected

Cartoon 6 Disruptive

What do you think of the behaviour of Child A ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child B ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child C ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child D ? 

Unacceptable — Better than expected

What do you think of the behaviour of Child E ? 

Unacceptable — Better than expected

Appendix 15

Teacher's Rating Scale of Children's Behaviours

Teacher's Rating Scale of children's behaviour

Rating example:

Could you please select 10 children randomly in your class and rate their behaviours according to the behaviours in the cartoons?

0 = Never 1= Occasionally 2= Often 3 = Very Often 4 = Almost Always

Name	Can't wait	Fidgety	inattentive	Disorganized	Distracted	Disruptive
James	0 1 2 ③ 4	0 ① 2 3 4	0 ① 2 3 4	0 1 2 ③ 4	0 1 ② 3 4	0 1 ② 3 4

0 = Never 1= Occasionally 2= Often 3 = Very Often 4 = Almost Always

Gender	Can't wait	Fidgety	inattentive	Disorganized	Distracted	Disruptive
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4

Appendix 16

Teacher's Rating Scale of Cartoon (in Chinese)

卡通评估教师

问卷教师背景信息 请对您选项□打√勾

学校: _____

性别: 女□ 男□ 班主任: 是□ 否□ 所教科目: _____

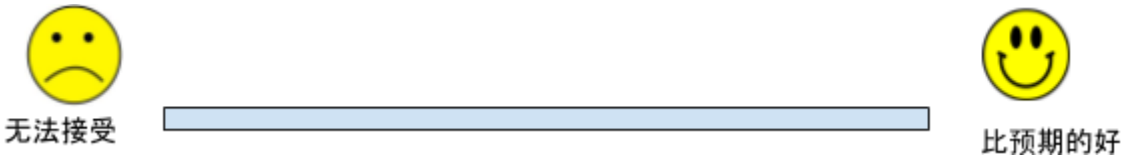
教龄: _____ 班级人数: _____

评估样例

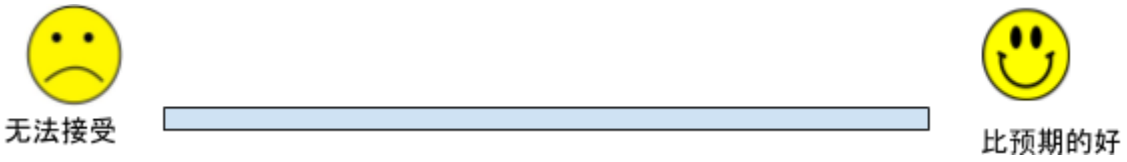
请您就学生A的行为在接受度上评分?

卡通-1 无法耐等待

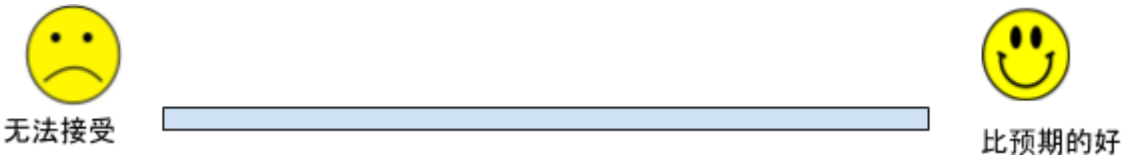
请您就学生A的行为在接受度上评分?



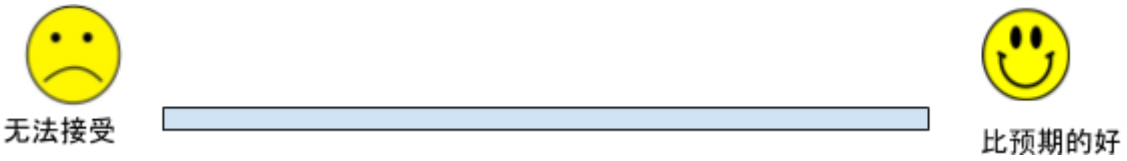
请您就学生B的行为在接受度上评分?



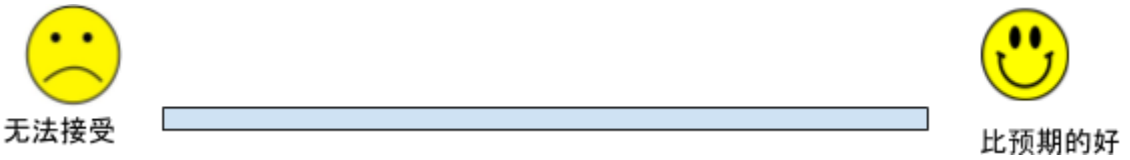
请您就学生C的行为在接受度上评分?



请您就学生D的行为在接受度上评分?

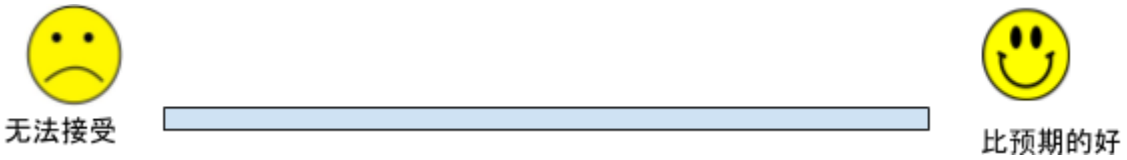


请您就学生E的行为在接受度上评分?

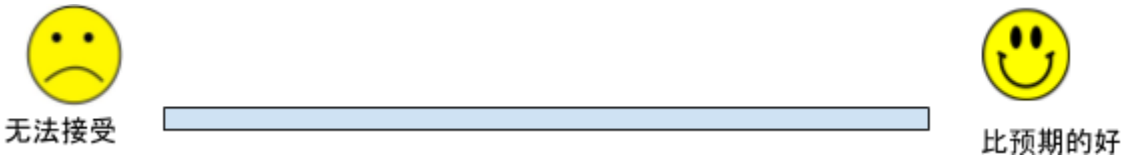


卡通-2 多动

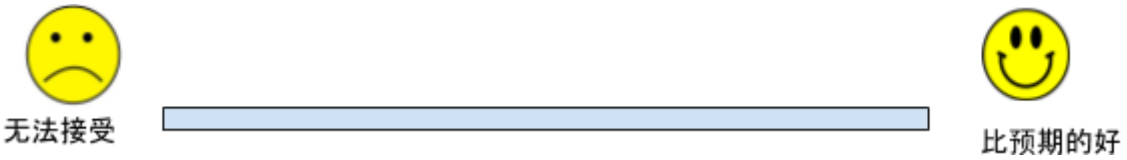
请您就学生A的行为在接受度上评分？



请您就学生B的行为在接受度上评分？



请您就学生C的行为在接受度上评分？

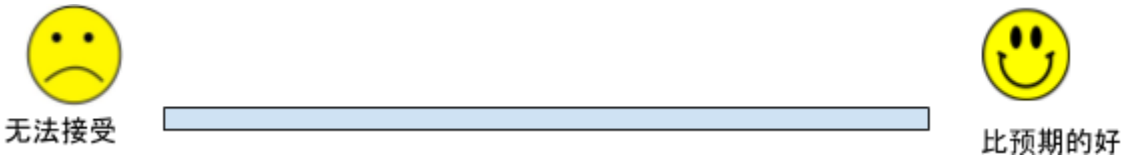


请您就学生D的行为在接受度上评分？

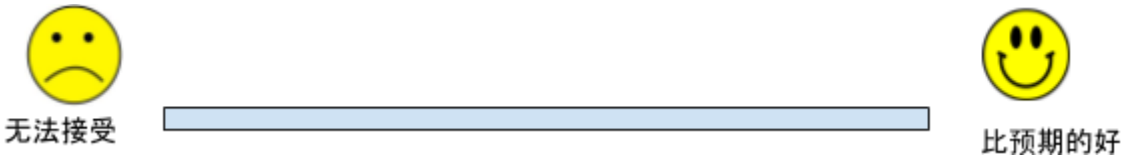
请您就学生E的行为在接受度上评分？

卡通-3 注意力不集中

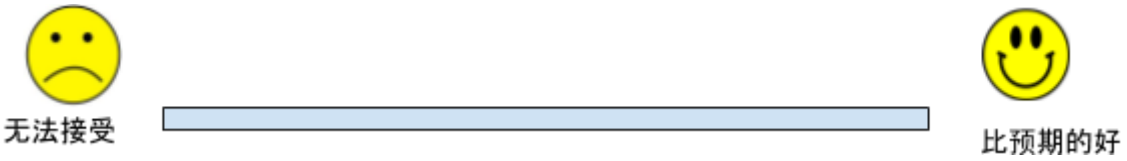
请您就学生A的行为在接受度上评分?



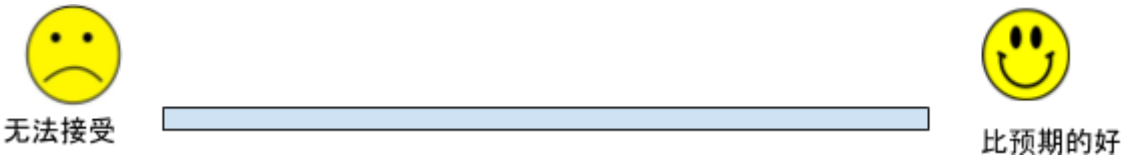
请您就学生B的行为在接受度上评分?



请您就学生C的行为在接受度上评分?



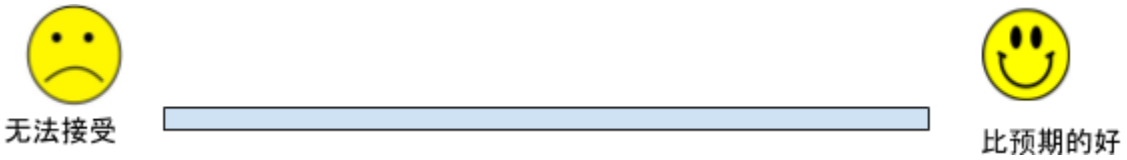
请您就学生D的行为在接受度上评分?



请您就学生E的行为在接受度上评分?

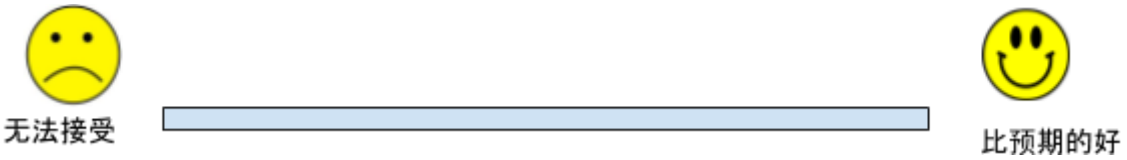
卡通-4 杂乱无章

请您就学生A的行为在接受度上评分？



请您就学生B的行为在接受度上评分？

请您就学生C的行为在接受度上评分？

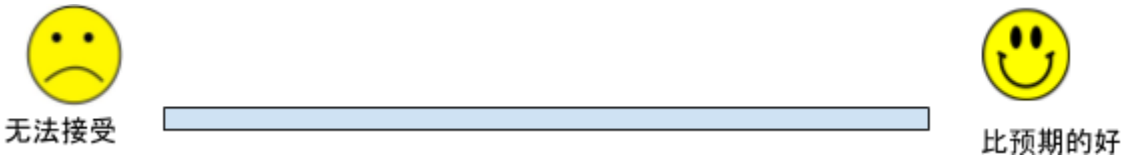


请您就学生D的行为在接受度上评分？

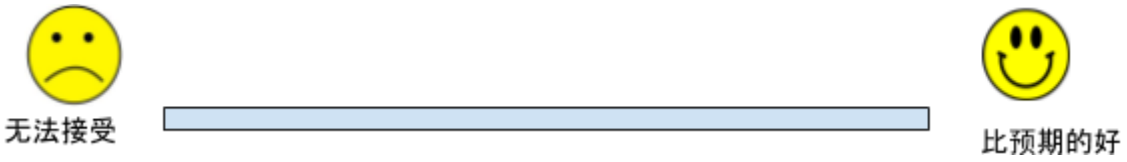
请您就学生E的行为在接受度上评分？

卡通-5 容易分心

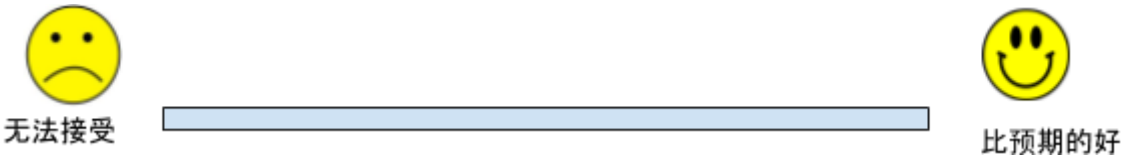
请您就学生A的行为在接受度上评分?



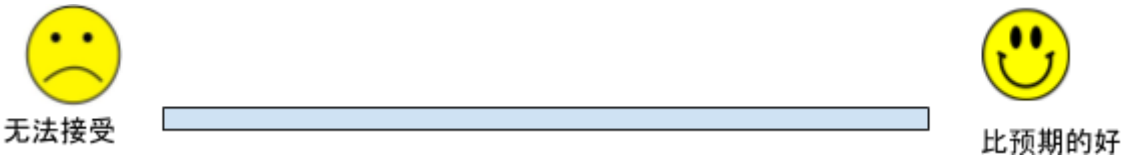
请您就学生B的行为在接受度上评分?



请您就学生C的行为在接受度上评分?



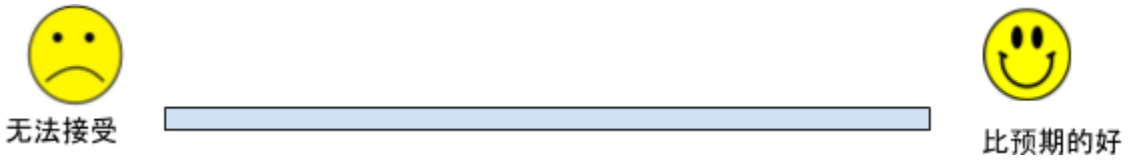
请您就学生D的行为在接受度上评分?



请您就学生E的行为在接受度上评分?

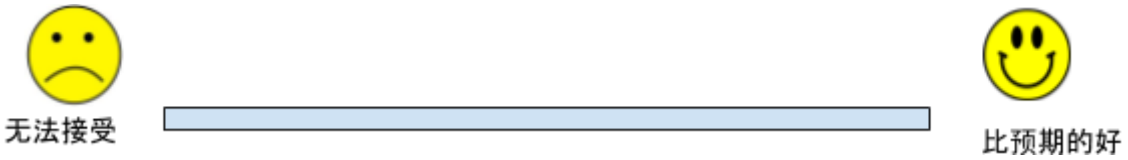
卡通-6 扰乱打扰别人

请您就学生A的行为在接受度上评分？



请您就学生B的行为在接受度上评分？

请您就学生C的行为在接受度上评分？



请您就学生D的行为在接受度上评分？

请您就学生E的行为在接受度上评分？

Appendix 17

Teacher's Rating Scale of Children's Behaviour (in Chinese)

儿童行为评估教师问卷

测评样例:

请您对您班级中随机挑选的10个学生的课堂行为根据卡通中儿童的行为进行评估

0 = 从不　1= 偶尔　2= 有时　3 = 経常　4 = 总是

姓名	无法耐心等待	多动	注意力不集中	杂乱无章	容易分心	扰乱打扰别人
王一	0 1 2 ③ 4	0 ① 2 3 4	0 ① 2 3 4	0 1 2 ③ 4	0 1 ② 3 4	0 1 ② 3 4

0 = 从不　1= 偶尔　2= 有时　3 = 経常　4 = 总是

姓名	无法耐心等待	多动	注意力不集中	杂乱无章	容易分心	扰乱打扰别人
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4

Appendix 18

Correlations between Teaching Years and English teachers' behavioural ratings

Behaviours		r	p (0.05, two-tailed)	n	Mean
A1	Leaving seat and asking for sweets	-.09	.68	23	3.83
B1	Shouting out for sweets	-.39	.07	23	5.09
C1	Waiting for	-.36	.09	23	8.57
D1	Leaving seat and eating sweets	-.04	.85	23	1.26
E1	Leaving seat and touching	-.21	.34	23	2.55
A2	Moving body on the seat	-.53*	.01	23	6.09
B2	Fiddling with a pencil and tapping feet	-.32	.14	23	5.96
C2	Moving chair	.19	.40	23	1.88
D2	Sitting still	-.26	.23	23	8.52
E2	Fiddling with other's hair	.25	.24	23	1.66
A3	Doodling	-.26	.22	23	1.98
B3	Tearing paper	.22	.32	23	1.07
C3	Daydreaming	-.25	.25	23	4.28
D3	Eating snack	.25	.25	23	1.64
E3	Concentrating	-.48*	.02	23	8.83
A4	Late for school	.22	.32	23	2.11
B4	Messy table	-.01	.70	23	4.19
C4	Tidy table	-.46*	.03	23	8.79
D4	Looking for things	-.03	.45	23	5.63
E4	Nothing on the table	-.32	.15	23	4.08
A5	Having a look and back to the lesson	-.25	.26	23	8.22
B5	Looking at the butterfly all the time	-.11	.61	23	8.12

C5	Playing with the butterfly but staying on seat	-.32	.14	23	3.63
D5	Shouting out	-.56**	.01	23	5.15
E5	Leaving seat to chaise the butterfly	-.33	.17	23	2.12
A6	Concentrating on writing	-.46*	.03	23	8.93
B6	Taking other's pencil without asking	-.27	.31	23	1.39
C6	Tickling another child	-.00	.97	23	2.25
D6	Doodling on other's writing	.21	.29	23	1.28
E6	Disturbing the teacher	.18	.40	23	4.76

Note. * Indicates p<.05, ** indicates p<.01

Appendix 19

Correlations between class size and English teachers' behavioural ratings

Behaviours		r	p (0.05, two-tailed)	n	Mean
A1	Leaving seat and asking for sweets	-.10	.64	23	3.83
B1	Shouting out for sweets	.15	.51	23	5.09
C1	Waiting for	-.01	.95	23	8.57
D1	Leaving seat and eating sweets	-.13	.56	23	1.26
E1	Leaving seat and touching	.25	.25	23	2.55
A2	Moving body on the seat	.11	.61	23	6.09
B2	Fiddling with a pencil and tapping feet	.00	.99	23	5.96
C2	Moving chair	-.38	.07	23	1.88
D2	Sitting still	.11	.61	23	8.52
E2	Fiddling with other's hair	-.16	.46	23	1.66
A3	Doodling	-.30	.16	23	1.98
B3	Tearing paper	-.24	.27	23	1.07
C3	Daydreaming	-.12	.59	23	4.28
D3	Eating snack	-.26	.24	23	1.64
E3	Concentrating	.15	.48	23	8.83
A4	Late for school	-.16	.47	23	2.11
B4	Messy table	.22	.32	23	4.19
C4	Tidy table	.12	.59	23	8.79
D4	Looking for things	.18	.41	23	5.63
E4	Nothing on the table	.04	.86	23	4.08
A5	Having a look and back to the lesson	-.15	.50	23	8.22
B5	Looking at the butterfly all the time	-.14	.54	23	8.12

C5	Playing with the butterfly but staying on seat	.18	.42	23	3.63
D5	Shouting out	.49*	.02	23	5.15
E5	Leaving seat to chaise the butterfly	.22	.32	23	2.12
A6	Concentrating on writing	-.11	.62	23	8.93
B6	Taking other's pencil without asking	.13	.55	23	1.39
C6	Tickling another child	-.13	.54	23	2.25
D6	Doodling on other's writing	-.21	.48	23	1.28
E6	Disturbing the teacher	-.33	.13	23	4.76

Note. * Indicates $p<.05$, ** indicates $p<.01$

Appendix 20

Correlations between Teaching Years and Chinese teachers' behavioural ratings

Behaviours		r	p (0.05, two-tailed)	n	Mean
A1	Leaving seat and asking for sweets	-.21*	.03	106	4.35
B1	Shouting out for sweets	-.17	.71	106	6.77
C1	Waiting for	.00	.89	106	7.88
D1	Leaving seat and eating sweets	.11	.64	106	2.33
E1	Leaving seat and touching	.24	.56	106	3.42
A2	Moving body on the seat	-.25	.68	106	3.96
B2	Fiddling with a pencil and tapping feet	-.31	.09	106	5.88
C2	Moving chair	.12	.47	106	2.08
D2	Sitting still	-.16	.82	106	8.17
E2	Fiddling with other's hair	-.23*	.02	106	3.02
A3	Doodling	-.28**	.00	106	2.43
B3	Tearing paper	.07	.78	106	2.65
C3	Daydreaming	-.22	.29	106	3.72
D3	Eating snack	.18	.62	106	2.9
E3	Concentrating	.13	.14	106	8.9
A4	Late for school	.04	.13	106	2.51
B4	Messy table	.33	.50	106	3.13
C4	Tidy table	-.18	.34	106	8.82
D4	Looking for things	-.25*	.01	106	3.49
E4	Nothing on the table	.30	.10	106	2.5
A5	Having a look and back to the lesson	-.14	.40	106	7.46
B5	Looking at the butterfly all the time	-.27	.09	106	6.7

C5	Playing with the butterfly but staying on seat	-.20	.26	106	2.35
D5	Shouting out	.10	.47	106	4.02
E5	Leaving seat to chaise the butterfly	.20	.61	106	1.33
A6	Concentrating on writing	.01	.84	106	8.51
B6	Taking other's pencil without asking	.02	.74	106	2.51
C6	Tickling another child	.19	.39	106	3.06
D6	Doodling on other's writing	-.15	.48	106	2.35
E6	Disturbing the teacher	-.29	.37	106	4.86

Note. * Indicates p<.05, ** indicates p<.01

Appendix 21

Correlations between class size and Chinese teachers' behavioural ratings

Behaviours		r	p (0.05, two-tailed)	n	Mean
A1	Leaving seat and asking for sweets	.18	.37	106	4.35
B1	Shouting out for sweets	-.00	.79	106	6.77
C1	Waiting for	-.32	.09	106	7.88
D1	Leaving seat and eating sweets	-.11	.53	106	2.33
E1	Leaving seat and touching	.22	.25	106	3.42
A2	Moving body on the seat	-.26	.25	106	3.96
B2	Fiddling with a pencil and tapping feet	-.24*	.01	106	5.88
C2	Moving chair	.25	.31	106	2.08
D2	Sitting still	.19	.45	106	8.17
E2	Fiddling with other's hair	-.21	.34	106	3.02
A3	Doodling	.36	.12	106	2.43
B3	Tearing paper	.04	.72	106	2.65
C3	Daydreaming	-.39	.08	106	3.72
D3	Eating snack	.13	.62	106	2.9
E3	Concentrating	.34	.25	106	8.9
A4	Late for school	-.15	.40	106	2.51
B4	Messy table	.25	.31	106	3.13
C4	Tidy table	.12	.61	106	8.82
D4	Looking for things	-.13	.54	106	3.49
E4	Nothing on the table	.16	.67	106	2.5
A5	Having a look and back to the lesson	.30	.10	106	7.46
B5	Looking at the butterfly all the time	-.27	.14	106	6.7

C5	Playing with the butterfly but staying on seat	.03	.25	106	2.35
D5	Shouting out	-.27	.74	106	4.02
E5	Leaving seat to chaise the butterfly	.31	.23	106	1.33
A6	Concentrating on writing	.10	.61	106	8.51
B6	Taking other's pencil without asking	.17	.46	106	2.51
C6	Tickling another child	-.21	.32	106	3.06
D6	Doodling on other's writing	.17	.60	106	2.35
E6	Disturbing the teacher	.09	.84	106	4.86

Note. * Indicates p<.05, ** indicates p<.01