Classification of Athletes with Intellectual Disabilities

Towards the re-inclusion of athletes with intellectual disabilities in the Paralympics

by Veronika van der Wardt

Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

30/09/2009

Abstract

The aim of this thesis was to investigate the relationship between intellectual functioning and sports performance for athletes with intellectual disabilities. A literature review showed that there is a significant difference between physical performance of athletes with and without intellectual disabilities, but so far, no studies have examined the association between the degree of intellectual functioning and sports performance for athletes with intellectual disabilities. Following an analysis of verbal and nonverbal intelligence tests on the basis of their psychometric properties, range of item difficulty, cultural fairness and duration of administration, the nonverbal SON-R 5½ - 17 was initially chosen to investigate the relationship between intellectual functioning and physical performance for athletes with intellectual disabilities. The findings revealed that this association depended on sports discipline: for table tennis, scores on a SON-R 5½ -17 subtest for inductive reasoning were associated with performance on the ABC physical aptitude test and sports competition performance (ABC physical aptitude test: R^2 adj.= 44%, beta= -.66, p<.01; table tennis competition performance: R^2 adj.= 30%, beta= -.66, p<.05), while for swimming and track and field athletics, none of the subtests (nor the overall IQ score) was significantly associated with physical or sports competition performance.

However, the results also revealed considerable limitations of this intelligence test for this research: All subtests showed floor-effects (zero scores) and comments from participants indicated that several items contained pictorial representations that were culturally biased.

Consequently, it was decided to develop a new computerized cognitive test battery for individuals with intellectual disabilities (CCIID), which was focused on the target population (individuals with intellectual disabilities) and based on theories of intelligence, research of cultural fairness, as well as the

2

results from previous testing using the SON-R 5½ -17 intelligence test. The test battery included two subtests for inductive reasoning and one subtest for visual processing abilities as these were the subtests shown to be associated with sports performance.

Psychometric properties of the CCIID were assessed for individuals with intellectual disabilities using modern and classical test theories. Based on the results of an item analysis using latent trait models and proportion of correctly scored items, several items were revised. A reliability study confirmed internal consistency (r = between .73 and .84, n = 60-66 depending on subtest, p < .05), test-retest reliability (r_s = between .77-.88, n = 24-27 depending on subtest, p < .05) and inter-rater reliability (r_s = between .42-.83, n = 22-25 depending on subtest, p < .05). An exploratory principal component analysis showed one underlying component with an Eigenvalue of 2.04, explaining 67% of the variance. This supported the construct validity of the CCIID. Criterion validity was confirmed based on correlations using Wechsler Adult Intelligence Scale scores (r_s= .66, p< .01, n= 18) and scores on the nonverbal SON-R 5½ - 17 intelligence test (r_s= .82, p< .01, n= 19).

Subsequently, the relationship between intellectual functioning and sports performance for athletes with intellectual disabilities was re-investigated using the CCIID. The results confirmed the findings of the initial studies: for table tennis, scores on the inductive reasoning subtest 'Series' significantly predicted table tennis performance (R^2 = .25, beta = .32, p<.05). Again, the studies revealed no significant association between scores on the CCIID and sports performance for swimmers and track athletes.

Further research should investigate if sports performance of swimmers and track and field athletes is limited through adaptive behaviour or different cognitive abilities using a wider range of cognitive information processing tests.
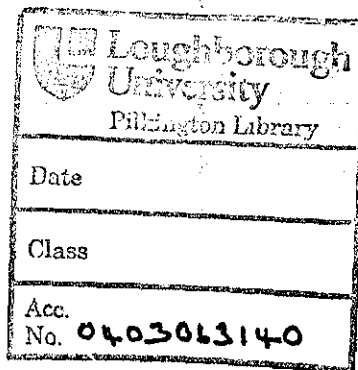
Acknowledgements

This study would not have been possible without the help of many people. First of all, I want to thank my supervisors Eef Hogervorst and Stephan Bandelow. I am very grateful for all the chances and challenges they gave me, their support, guidance and advice. I could not have had better supervisors. Thanks for giving me great opportunities to start my way as a researcher.

I thank Ken Black for getting the project started, and Herman van Boxtel, my Masters supervisor in Utrecht, for his advice and support at the very start of the project.

I want to thank all the participants who volunteered to take part in this research, especially the British Table Tennis team who have been fantastically patient with our repeated appearances, and their parents, carers and coaches who were always cooperative. Then I want to thank INAS-FID and MENCAP for their support, in particular Nick Parr and Bob Price for facilitating testing opportunities. I would also like to express my gratitude to all the people who helped testing, the ID research group for their constructive comments when I presented our research, all the schools and organisations who enabled the access to their students and members, the organisers of the 2008 European and Open Table Tennis Championships in Ontinyent/Spain and the organisers of the 2009 Global Games in Liberec/Czech Republic for supporting our research during their events.

And, of course, I am very grateful to my family, my children Ben and Elena who saw me mostly sitting behind the computer in these last weeks, Derk, for his love and confidence in me, my children's grandparents who came many times to help out and my sister Irena who was a great help in the summer holidays. I also want to thank all my friends who had to listen to endless

stories about my work in the last months and who still invited me around for coffees, cinema and camping.

Finally, I would like to thank everybody who was important to the successful realisation of the thesis, as well as expressing my apologies to everybody I could not mention personally.

8

9

Index of tables

12

Chapter 1: Introduction

Cognitive abilities are an essential factor in elite sports performance (Davis & Sime, 2005; Thomas, 1994). Research investigating the nature of this relationship is limited, in particular for athletes with intellectual disabilities (ID), but is essential for elite sports development for this group in several ways, which will be discussed later in this chapter.

The purpose of this thesis was to investigate the relationship between cognitive abilities and sports performance for elite athletes with ID. To examine this relationship, scores on a nonverbal intelligence test were compared to physical as well as sports performance. Although the results of these studies confirmed the association, limitations of the intelligence test also showed the need for a new instrument, specifically developed for individuals with ID. This led to the development of the Computerized Cognitive test battery for Individuals with Intellectual Disabilities (CCIID). The design of the CCIID was based on theoretical considerations of intelligence theories, cultural fairness, as well as the results of the initial studies. Psychometric properties of the CCIID were assessed using classical and modern test theories. Subsequently, the association between cognitive abilities and sports performance in athletes with ID was re-investigated using the new test battery.

Even for non-disabled athletes, research investigating the association between cognitive abilities and sports performance is not comprehensive and has only examined single sports disciplines. Although the importance of optimal cognitive function in elite sports performance is generally accepted, the investigation of the relationship is limited to particular cognitive abilities and sports disciplines. In these studies, the term 'cognitive abilities' refers to a

large variety of mental processes, such as perception, memory, attention, problem solving, information processing, reasoning and concept formation.

Previous studies suggested that, for non-disabled athletes, different cognitive abilities were relevant, depending on sports discipline: for instance, elite rugby players showed significantly higher scores on cognitive tests for visuo-spatial abilities and processing speed than controls (Kasahara, Mashiko & Niwa, 2008). Ryan, Atkinson and Dunham (2004) compared different sports disciplines using several cognitive ability measures, and found that hockey players performed significantly higher on perceptual-motor and accuracy tasks than swimmers or track and field athletes. This difference between sports disciplines in cognitive performance was also supported by the findings of Overney, Blanke and Herzog (2008) who demonstrated that elite table tennis players showed significant better performances in visual discrimination tasks requiring different visual aspects, such as motion detection, attention and temporal processing than elite tri-athletes and non-athletes. Decision-making was also found to differ between sports disciplines. A comparative study including tennis players, table tennis players, fencers and boxers competing on a national level showed that tennis players and fencers demonstrated faster and more accurate responses in a choice reaction time task than table tennis players and boxers (Mouelhi Guizani, Tenenbaum, Bouzaouach, Ben Kheder, Feki & Bouaziz, 2006).

Furthermore, various studies found that experts or elite athletes performed significantly better than novices in various cognitive abilities, but again, this depended on sports discipline. Helsen and Pauwels (1993) showed that expert soccer players responded faster and more accurately in a sport-specific visual information processing task than novices. Similar results were found in an experimental study, which demonstrated that expert tennis players responded significantly faster and more accurately to visual cues than novices (Shim, Carlton, Chow & Chae, 2005). Experts also displayed better

performances than novices in making the appropriate decisions and selecting the optimal responses in a sport-specific context (De Villar, Gonzalez, Iglesias, Perla & Cervello, 2007; Nougier, Stein & Bonnel, 1991). Several cognitive abilities were investigated in a meta-analysis (Mann, Williams, Ward & Janelle, 2007), which revealed that sport experts demonstrated better response times and accuracy than novices, showed fewer visual fixations suggesting they allocated less resources to information processing and displayed longer 'quiet eye' episodes. This indicated that task-relevant information is better processed and motor plans are better coordinated in expert athletes.

Although these studies indicated that experts demonstrated superior cognitive abilities than novices, it is not clear, if and to what extent these abilities are transferable to a context outside the expertise of the athlete. In addition, the causality of the relationship between cognitive abilities and sports performance is debatable: Did athletes learn these cognitive abilities during their training, or did those athletes with lower abilities drop out of their sporting career?

Independent of the causality, intellectual disabilities are likely to affect the development of elite sports performance for disciplines with a high relevance of cognitive abilities. It can be expected that athletes with ID will not be able to compete on the same level as athletes without disabilities in these disciplines. However, the relationship between cognitive abilities and sport performance has been insufficiently examined for individuals with ID to support this conclusion empirically.

Research regarding this relationship is important to advance the elite sport of individuals with ID on different levels:
- Separate sports competitions for individuals with and without ID are only necessary if they cannot reach the same level of performance due

15

to their disabilities. Currently, athletes with ID are not allowed to participate in the Paralympics until the impact of intellectual disability on sports performance is actually confirmed. After the 2000 Sydney Paralympics, several Spanish gold-medalist basketball players were caught cheating as they pretended to have intellectual disabilities. Consequently, the International Paralympic Committee (IPC) decided to suspend the category 'Intellectual Disability' from the Paralympics. As a result, athletes with ID cannot enter competitions in the Paralympics until a) the impact of ID on sports performance is actually confirmed and b) a classification system is found that can confirm the impairment in those cognitive abilities that affect the sports performance in the discipline of the athlete (IPC, 2007). Both requirements depend on the determination of the relationship between cognitive abilities and sports performance of athletes with ID.

- Talent identification of athletes with disabilities is still in its early stages. Research into the relationship between cognitive abilities and sports performance for athletes with ID could identify those cognitive abilities related to superior sports performance in different disciplines. Further research could then establish if these cognitive abilities are predictive for future sports performances to identify talented young athletes.

Prior to participation in sport events, the disability of athletes with ID has to have been established. According to the World Health Organisation (WHO, 2007) and the American Psychiatric Association (APA, 2000), intellectual disability is defined by three criteria: 1) onset of the disability before the age of 18; 2) impairment of intellectual functioning and 3) significant limitations in adaptive behaviour. Impairment of intellectual functioning is commonly measured using a standardized intelligence test, while limitations in adaptive behaviour are assessed using adequate scales. The two criteria, impairment of intellectual functioning and limitations in adaptive behaviour, determine the degree of the disability: The lower the IQ score and the more limitations in

16

adaptive behaviour, the higher the degree of the intellectual disability. Therefore, IQ score as well as limitations in adaptive behaviour will have to be investigated to determine the association between intellectual disability and sports performance.

The British term for intellectual disability is 'Learning Disability' which is used synonymously in British scientific literature (Cooper, Melville & Morrison, 2004; Cornwell, 2004). For the sake of consistency, the term 'intellectual disability' is used throughout the thesis. The British guidance for 'learning disability' within the British Psychological Society refers to the same international standards as the World Health Organisation or the DSM-IV (intellectual impairment, limitations in adaptive behaviour and onset before the age of 18) but does not necessarily base the limitations of intellectual functioning on an IQ score, but on the judgment of a psychologist. Not all British athletes with learning disabilities are diagnosed with an IQ test. However, in order to establish their eligibility for international sports events in the category 'Intellectual Disability', they do need to provide the results of an IQ assessment including the scores on an intelligence test.

INAS-FID is the International Federation for sport for athletes with an intellectual disability. INAS-FID organizes sport events and is responsible for the registration of athletes with ID for sport events which could potentially lead up to the Paralympics (depending on the decision about inclusion of athletes with ID). Athletes can register and compete for INAS-FID events if they can provide evidence for an IQ score of 75 or below, significant limitations in adaptive behaviour and an onset of the disability before the age of 18.

The studies of this thesis will be limited to the cognitive aspects of intellectual disability. The association between adaptive behaviour and sports performance of athletes with ID is outside the focus of this study and will, therefore, not be discussed.

17

The following sections will discuss physical performance of individuals with ID and examine different intelligence tests, which could be used to investigate the association between intellectual functioning and physical as well as sport performance in individuals with ID.

There are several explanations of how cognitive impairments of athletes with ID could limit physical performance, as well as sport performance. These will be discussed in the following section.

Physical performance of individuals with ID has been widely researched over the last 15 years. Most studies focused on the differences between sedentary individuals with and without intellectual disabilities, and only a few studies investigated athletes with ID. A literature review, presented in chapter 1, section 1.1.2, evaluated existing knowledge about physical performance of individuals with ID and identified limitations of these studies.

Different intelligence tests were analysed as the investigation of the relationship between intellectual impairment and sports performance required an intelligence test suitable for the target group (athletes with ID from different countries). A review of intelligence tests in chapter 1, section 1.2. examined the requirements for an appropriate test, discussed psychometric issues and looked at different intelligence tests. On the basis of this analysis, a suitable intelligence test was selected, which was subsequently used to examine the association between cognitive abilities and physical as well as sports performance in athletes with ID. This relationship was investigated for different sports disciplines as it is likely to vary between sports disciplines.

1.1.1. Possible causes for the impact of intellectual disability on physical performance

Before examining the differences in physical performance between individuals with and without intellectual disabilities in the next section, the underlying biological causes for these differences will be investigated. Up to now,

research has identified several biological causes that underlie physical performance and intellectual disability:

1. Genetic disorders, such as Down's syndrome, William's syndrome or fragile x-syndrome can be a common cause for the association between intellectual disability and physical performance (Black, Smith, Wu & Ulrich, 2007; Charlton, Ihsen & Lavelle, 2000; Hagerman & Hagerman, 2002; Kubo & Ulrich, 2006; Morris & Mervis, 1999). The symptoms of these genetic disorders include impaired cognitive functioning, as well as physical disabilities. Even if the intellectual disability is not caused by a genetic disorder, other biological reasons can cause the association between physical performance and intellectual disability, such as:

2. Pre-term birth (before 33 weeks' gestation) can lead to brain lesions and a subsequently reduced cerebellar volume (Nguyen The Tich, Anderson, Shimony, Hunt, Doyle & Inder, 2009). This can cause lasting cognitive impairments and motor function problems (Allin & al., 2001; Hall, McLeod, Counsell, Thomson & Mutch, 1995). Brain lesions and impairment of the cerebellum due to very pre-term birth might, therefore, be a possible cause for impairments in intellectual functioning, as well as in physical performance.

3. Another possible cause for intellectual disability is presented by white matter hyperintensities (WMH), which are frequently associated with cognitive impairments (Gunning-Dixon & Raz, 2000) and physical deficits, such as gait and balance problems (Steffens, Bosworth, Provenzale & al., 2002), fine motor coordination and grip strength (Sachdev, Wen, Christensen & Jorm, 2004). Although WMH are more common in an elderly population, children diagnosed with WHM present similar cognitive and motor function impairments (Tartaglia & al., 2008). White matter abnormalities are also found in individuals with developmental delay (Widjaja, Nilsson, Blaser & Raybaud, 2008) and cerebral palsy (Robinson, Peake, Ditchfield, Reid, Lanigan &

Reddhough, 2008), which again are frequently linked to physical, as well as cognitive disabilities (Dinnage, 1986; Belligni & al., 2009).

4. Individuals with ID often have additional physical disabilities (Eichstaedt & Lavay, 1992). For instance, Wuang, Wang, Huang and Su (2008) found that scores on the Wechsler Intelligence Scale for Children were positively associated with fine and gross motor skills in early school-age children with mild ID. A comparison showed that individuals with ID had significantly lower perceptual-motor coordination than individuals without ID (Carmeli, Bar-Yossef, Ariav, Levy & Livermann, 2008). In a study investigating balance and coordination, the results showed significant lower scores in different sensorimotor tests in adults with ID than in non-disabled controls (Carmeli, Bar-Yosssef, Ariav, Paz, Sabbag & Levy, 2008). A comparative study including individuals with and without ID showed that strength measures and endurance are lower in individuals with ID than in non-disabled controls (Lahtinen, Rintala & Malin, 2007). In an investigation assessing grip strength during the Texas Special Olympics, O'Connell, Rutland and O'Connell (2006) found that individuals with ID had significantly lower grip strength than the age-matched controls.

Limitations in sports performance cannot only stem from common causes underlying both physical, as well as cognitive impairments, but could also be the direct result of the intellectual disability. Cognitive impairment will affect the ability to understand, memorize and transfer instructions given by sports coaches. These instructions form an important part of the skills acquisition in sports (Allison & Ayllon, 1980). Hodges and Franks (2002) reported that the selection and execution of instructions and demonstrations given by the coach requires attention and cognitive processing, particularly in the early stage of skill acquisition. Implicit and explicit motor learning was investigated in a comparison between children with and without Down syndrome. Although both groups performed equally well in the implicit learning condition, the

results showed that children with Down syndrome performed at a significant lower level than non-disabled controls in the explicit learning condition (Vinter & Detable, 2008). The extent to which the intellectual disability influences the skill acquisition will depend on the degree and nature of the disability (Horvat, 1990). Consequently, it can be concluded that sports performance is influenced by cognitive functioning, which subsequently affects skills acquisition.

Despite the fact that intellectual disability is often linked to limitations in physical performance, training was found to improve physical performance of individuals with ID (Mached, Stopka, Tillman, Sneed & Naugle, 2008, Tsimares& Fotiadou, 2004). Therefore, it is necessary to investigate not only the causes underlying the association between intellectual disability and physical performance, but the actual impact of intellectual disability on physical performance itself, which has been reported in many studies. The following section will look at studies examining the difference in physical performance between individuals with and without ID and the limitations of these studies.

1.1.2. Research on physical performance of individuals with intellectual disabilities

Various studies have investigated fitness parameters such as cardiovascular fitness, muscular strengths and obesity. The results indicated that individuals with ID are physically less fit than non-disabled individuals:

Cardiovascular fitness (CVF) was investigated in two reviews (Fernhall, Tymeson & Webster, 1988, Lavay, Reid & Cressler-Chaviz, 1990), which both reported that most studies found lower than average performances on CVF parameters for individuals with ID. CVF observed in children and adolescents with ID, when measured through field walk-run tests or 300-yard runs, were around 25-30% lower than those of children and

adolescents without ID (Rarick, Widdop & Broadhead, 1970; Halle, Silverman & Regan, 1983). CVF levels in adults were measured in walk-run tests or a 1.5 mile run and were also substantially below those of adults in the general population (Cressler, Lavay & Giese, 1988; Pitetti & Campell, 1991).

In addition, studies reported that individuals with ID had lower muscular strength and endurance than individuals without disabilities. For instance, the results of field tests with children with ID showed below average sit-up and pull-up performances, compared to non-disabled peers (Rarick & al., 1970). Investigations based on adults with ID showed similar results: muscular strength and endurance measured in sit-ups and push-ups were significantly below average, with men performing significantly worse than women (Reid, Montgomery & Seidl, 1985).

Although findings suggest a prevalence of obesity for individuals with ID when diagnosed with mild or moderate intellectual disability, this is not the case for individuals diagnosed with severe or profound intellectual disability (Hove, 2004). For children, the difference in prevalence in obesity between intellectually disabled and non-disabled groups is reported to be relatively small (Maksud & Hamilton, 1975) or insignificant (Pizarro, 1990). For adults, obesity has a higher prevalence in intellectually disabled individuals (when diagnosed with mild or moderate ID) compared to non-disabled individuals (Fox, Burkhart & Rotatori, 1983). Therefore, it can be assumed that obesity is only a minor factor in the difference in physical performance between children with and without ID, while it is a more important aspect in the difference in physical performance between adults with and without ID.

In sum, all these studies suggest that individuals with ID have an inferior physical performance compared to their non-disabled peers. This difference could be due to the causes that underlie both intellectual disability and physical performance. However, as Fernhall (1993) and Pitetti and Campell (1991) noted, various other factors have been established to be responsible for the inferior physical performance of individuals with ID. These factors are described below:

22

First, physical performance can be influenced by lifestyle, e.g. physical exercise, smoking, drinking, dietary habits, etc. A study evaluating the effect of a 16-week training program indicated that physical exercise can have a substantial influence on cardiovascular fitness (CVF) for individuals with ID (Pitetti & Tan, 1990; 1991). Similar results for improvements in CVF were found in other training studies (Tomporowski & Ellis, 1985; Croce, 1990). Draheim et al. (2002) investigated differences between trained and untrained individuals with ID in cardiovascular disease risk factors. Their results showed that trained individuals with ID (participants of the Special Olympics) had significant lower diastolic blood pressures, body fat percentages, abdominal fat, triglycerides and insulin than inactive individuals with ID. Consequently, it can be concluded that exercise has a significant influence on physical performance for intellectual disabled individuals.

In addition, other lifestyle factors, such as dietary habits, smoking and drinking are regarded to have the same impact on individuals with ID as they have on the general population (Van de Louw, Vorstenbosch, Vinck, Penning & Evenhuis, 2009; Wallace & Schluter, 2008) and, therefore, will contribute to the physical performance of individuals with ID. Consequently, studies should control for these lifestyle factors.

Second, physical performance, when measured in scientific studies, could be influenced by a lack of motivation. Motivation is particularly important for studies, which require the participant to perform with maximum effort. Studies suggested that individuals with ID interpret instructions like "give your best effort" differently to individuals without disabilities as they might stop when feeling breathlessness, lack the ability to pace themselves over an extended period of time, or lack the persistence to continue beyond feeling fatigued which can be necessary to measure fitness parameters such as CVF (Baumgartner & Horvat, 1988; Lavay & al., 1990). Furthermore, motivational problems seem to expand with increasing levels of intellectual disability (Fait & Dunn, 1984).

Third, although most of the measures for physical performance are proven valid and reliable for the general population, they are not investigated for their validity and reliability to test individuals with ID (Pitetti & Tan, 1990; Lavay & al., 1990).

In addition, other possible confounding factors should be considered which have not yet been included in any of the studies. Possible factors could be the range of IQ scores of the participants, differences in hours of training and coaching, differences in training methods, etc. These factors should all be investigated in future studies.

It is important to note that, although there are differences in physical performance between individuals with and without ID, these differences can be caused by the disability itself, or confounding factors. The following studies compared physical performance of individuals with and without ID, but controlled for some of these factors.

In a retrospective study, Fernhall et al. (1996) collected records of 111 individuals with ID in the USA. These records contained cardiorespiratory data (VO2 peak[1], VE peak[2], peak heart rate and peak respiratory exchange ratio) collected during treadmill exercises in a laboratory setting. All participants had been familiarized with the locations and the exercises. The testing protocol had been proven valid and reliable for individuals with ID (Fernhall, Millar, Tymeson & Burkett, 1991). Results were controlled for age, height, weight and gender and showed that individuals with ID had sub-average cardiorespiratory levels compared to the general population.

Although this study used measures that are proven valid and reliable, the results were not controlled for lifestyle (as weight alone is not an accepted lifestyle indicator) and motivation. In addition, it is not clear if participants

---

[1] VO2 peak is a plateau effect in oxygen consumption, while workload continues to increase
[2] VE peak is the total volume of gas in liters exhaled from the lungs per minute at peak exercise

completed the task with maximum effort and if, therefore, the highest possible values for cardiorespiratory data were obtained (Fernhall & al., 1996).

In a study assessing flexibility and strength, Pitetti and Yarmer (2002) compared 269 children and adolescents without ID with 449 children and adolescents with ID (diagnosed with mild mental retardation) in knee flexion, knee extension and combined leg and back strength measures. Each measurement was taken twice and the better result used for the analyses. The instrument to measure knee flexion and extension was proven to be valid and reliable for children and adolescents with and without ID (Croce, Horvat & Pitetti, 1999; Hill, Croce, Miller & Cleland, 1996). However, validity and reliability had not been proven for the instrument assessing combined leg and back strength. The results showed that individuals with ID had significantly lower strengths levels in all age groups (8 to 10 years, 11 to 14 years and 15 to 18 years) compared to their non-disabled peers when controlled for sex.

Again, the study indicated that there is a significant difference in physical performance between individuals with and without ID. However, only one of the two instruments was proven valid and reliable for the use for individuals with ID but correlation coefficients between the first and the second measurement for all participants were significantly high (Pearson's r between .82 and .95, depending on test). Again, differences in levels of exercise, lifestyle and motivation between individuals with and without ID might have influenced the outcomes.

In another study, Pitetti, Yarmer and Fernhall (2001) compared aerobic fitness and body mass index (BMI) of children and adolescents with and without mild ID. Aerobic fitness was measured with the 20 meter shuttle run (20-MST). The 20-MST was proven to be a valid and reliable instrument to assess aerobic fitness for children and adolescents (Leger, Mercier, Gadoury & Lambert, 1988; Fernhall & al., 1998). Independent of age and sex, individuals without ID ran significantly more laps and had a lower BMI than

25

their peers with ID. When the influence of BMI on the aerobic test was controlled for, the individuals with ID still showed lower aerobic fitness.

A similar study was conducted by Pitetti and Fernhall (2004) who also compared running performance in a 20-MST of youth (11-18 years old) with and without ID. All participants received verbal encouragement during the run. Again, results showed, after controlling for age, BMI and sex, that individuals without ID performed better than their peers with ID.

An investigation with the Eurofit Special test battery (Skowronski, Horvat, Nocera, Roswal & Croce, 2009) compared three groups of individuals with ID (mild, moderate and severe ID). The Eurofit Special measured strength, local muscle endurance, speed, flexibility and balance. Reliability of the Eurofit test battery was established for individuals with ID (MacDonncha, Watson, McSweeney & O'Donovan, 1999). The results showed a significant difference in performance between individuals with mild, moderate and severe ID. This study indicated a relationship between intellectual disability and physical performance, but did not control for motivation and lifestyle issues, which might have influenced the outcome.

None of these studies controlled for all possible confounding factors, but they excluded many of them: for most studies, instruments were proven valid and reliable, and the results were controlled for BMI. However, other lifestyle factors such as alcohol consumption, smoking and dietary habits were not included, and a difference in motivation between individuals with and without intellectual disability might have further influenced the results.

Additionally, it is interesting to note that the studies from Pitetti, Yarmer and Fernhall (2001), Pitetti and Yarmer (2002) and Pitetti and Fernhall (2004) all described their participants as being recruited in a summer camp in a Midwest metropolitan area in the USA. Therefore, it is possible that these three studies recruited the same participants or from the same pool of participants making a generalization of results difficult.

Consequently, it would be necessary to compare physical performance of individuals with and without ID with valid and reliable instruments while controlling for lifestyle factors and motivation in order to confirm the results of these studies.

1.1.3. Research on physical performance of athletes with intellectual disabilities

In a different approach to tackle the problem of controlling for lifestyle, motivation and other confounding factors, the following two studies chose athletes who competed in international sports events as participants. The assumptions were that elite athletes would have a healthy lifestyle, be accustomed to presenting a maximum performance in order to be successful in their competitions and would have similar training hours, methods and coaching to non-disabled elite athletes.

The first study investigated physical fitness of runners with and without ID (Frey & al., 1999). The nine participants with ID were qualified for the Special Olympics and were involved in a running program. The non-disabled participants were matched for age, gender and weekly training hours. The measurements taken were cardio-respiratory fitness (with a treadmill), leg strengths (using a computer assisted isokinetic dynamometry), flexibility (sit and reach test) and percentage body fat. Except for the cardio-respiratory fitness exercise (Fernhall, Millar, Tymeson & Burkett, 1991), none of the other instruments was proven to be valid or reliable for individuals with ID. The results showed no significant difference between the groups with and without intellectual disability in cardio-respiratory fitness, percentage body fat or flexibility. However, the results did reveal a significant difference between the groups in leg strength. Runners without ID demonstrated significant greater knee flexion and extension peak torques compared to runners with ID.

Although the study did find significant differences in leg strength, it has to be noted that the sample was very small (9 participants with ID), two of the

three instruments did not have proven reliability or validity for individuals with ID and, although body fat measures were taken, the results of the fitness measures were not controlled for that variable.

The second study compared physical fitness of elite athletes with and without ID (Van de Vliet & al., 2006). The 776 participants with ID were athletes who took part in the 2004 Global Games in Bollnäs, Sweden, where they were also tested. Physical performance of the athletes with ID was established using the EUROFIT test battery (European test of physical fitness). This test battery measures whole body balance, speed of limb movement, flexibility, explosive strength, static strength, abdominal muscular endurance, upper body muscular endurance and running speed. The EUROFIT test has been proven reliable for male adolescents with and without ID (MacDonncha, Watson, McSweeney & O'Donovan, 1999). Validity of the EUROFIT had not been established for individuals with ID. Additionally, age, height, weight, percentage body fat and body mass index (BMI) were assessed. The results were compared with physical fitness data from three different studies with non-disabled university sports students. Although sports students were physically very active, they were not training at the high level of the intellectually disabled participants of the Global Games. The comparison indicated that both male and female athletes with ID demonstrated lower fitness levels for cardio-respiratory endurance, speed of limb movement, explosive strength, abdominal muscular endurance and hand grip strength. For all other measurements the athletes with ID had similar or higher fitness levels compared to the non-disabled sports students.

This study indicated that there are significant differences in physical performance between individuals with and without ID. Differences between the groups due to motivation and lifestyle were controlled for by choosing elite athletes as participants, and the instrument was proven to be reliable (although not proven to be valid).

1.1.4. Interim conclusion

In summary, the results of all different studies suggest that there is indeed a significant difference in physical performance between individuals with and without ID. Although some studies did not control for all possibly confounding factors, and not all studies found significant differences in all physical performance parameters, none of the results found individuals with ID performed better than individuals without ID when their physical activity levels were matched. Therefore, an association between physical performance and intellectual disability can be concluded. However, it is important to note that research concerning physical performance so far is limited to comparisons between individuals with and without ID. None of the studies had compared the level of physical performance for different degrees of ID. Such a comparison would be essential for the development of a classification system for sports competitions for athletes with ID, which depends on the degree of function in physical performance as it is presented by the disability. Consequently, it is necessary to determine if the association between physical performance and intellectual disability is equally present for different degrees of intellectual disability.

As discussed earlier, elite sports performance does not only depend on physical performance, but also on different cognitive abilities. However, the relationship between sports performance and cognitive abilities has not yet been investigated for athletes with ID. Although it is likely that intellectual disability limits sports performance in various domains such as instructions (comprehension of instructions, transfer of instructions to different situations), visual perception, self-discipline, attention, mental rehearsal, self-efficacy and motivation (Ericsson & Lehmann, 1996; Kane, 1979), intellectual functioning has not been investigated in relation to sports performance of athletes with ID.

The following section will establish a suitable intelligence test to examine the relationship between intellectual functioning and sports performance of athletes with ID.

1.2. Intelligence tests for individuals with intellectual disabilities

1.2.1. Requirements for the intelligence test

Any test used should be suitable for individuals with ID. Although most intelligence tests will identify an IQ under 75, many tests are less reliable and accurate towards the lower and higher ends of the intelligence range (Nunnally & Bernstein, 1994). The fact that most participants competing in sport events for athletes with ID are teenagers and young adults, complicates matters: Most intelligence tests cover either the age range of teenagers (intelligence tests for children) or the age range of young adults (intelligence tests for adults). This point will be discussed in greater length in section 1.2.2, but, at this point, it is important to note that any test will have to be examined for its appropriateness for teenagers and young adults with ID.

Eligibility determines if an athlete can participate in a sports competition for athletes with disabilities within a certain category of disability. Most categories are defined by an obvious physical disability and, therefore, do not need a verification test of the disability. Intellectual disability is often less obvious. Intelligence tests so far cannot distinguish between individuals who do have ID and individuals who *pretend* to have ID. Until a method is found that can establish if a person cheats, possibilities to cheat should be minimized in order to allow athletes with ID to compete in the Paralympics. A short intelligence test could be used and supervised on site of the sporting event. Although that would not eliminate the possibility that a participant consciously tries to score lower on the intelligence test, it would ensure that

each participant goes through the same testing procedure. Therefore, an intelligence test of short duration is preferable.

Participants of the Paralympics come from all over the world. The intelligence test used to determine their eligibility and classification, thus has to be cross-cultural insensitive. Although it is unlikely that any test is entirely insensitive to the cultural background (Sattler, 1992), cultural fairness was discussed for all tests included in the following review in section 1.2.3.

Consequently, five criteria must be examined: First, the suitability of the intelligence test for intellectual disabled individuals; second, the aptness of the age range of the test for teenagers and young adults with ID; third, psychometric properties; fourth, the duration of the administration of the test; and last, cross-cultural sensitivity. Each of the intelligence tests presented in section 1.2.3. was studied on the basis of these criteria.

1.2.2. Psychometric issues

In order to evaluate an intelligence test for its use for individuals with ID, it is necessary to discuss several psychometric concerns:

Depending on the purpose for which the intelligence test was constructed, an intelligence test will contain a certain number of items of different levels of difficulty. For the purpose of this study, it is particularly important that the test has a sufficient number of items at different easy levels in order to prevent a 'floor' effect, where even the easiest items are too difficult for the target population. In order to be able to get meaningful results, it is essential that the test provides enough items on different easy levels to differentiate between their IQ scores.

Generally speaking, intelligence tests are less reliable at the ends of the score range than in the middle of that range (Nunnally & Bernstein, 1994). Therefore, a test chosen to assess individuals with ID will only be useful if it also discriminates reliably at the lower end of the score distribution. A test constructed as an equidiscriminating (EQD) test could minimize that problem,

31

although, in order to increase the reliability at the ends of the range (highest and lowest percentiles), the test would be required to include a high number of items. Then again, the long duration of such a test could potentially lead to other problems when testing a population with ID (e.g. attention problems or loss of motivation).

Another concern is the transformation of raw scores into standardized scores. Depending on the transformation process, differences in raw scores near the mean are enlarged when transformed into standardized scores and reduced in the tails of the distribution (Cronbach, 1990). Intellectually disabled individuals will usually obtain low raw scores. The important differences in these low scores will get lost when the raw score is transformed into a deviation IQ score which is common for intelligence tests.

A further issue is the age range for which the different intelligence tests are designed. Most intelligence tests are constructed either for children or adults but do not cover teenagers and young adults. There are exceptions though, like the Kaufman Adolescent and Adult Intelligence Test (KAIT) which can be used from an age of eleven years onwards.

In summary, there are several psychometric issues regarding intelligence tests used for the assessment of individuals with ID as their scores are usually in the lower range of the distribution of the scores. Using tests, which include easier items would lead to higher scores and a better differentiation for individuals with ID. Furthermore, it is likely that 'floor' effects could be reduced if not avoided. Consequently, it was decided to review intelligence tests for this project which were designed for children and teenagers. It can be expected that teenagers and young adults with ID score higher on tests constructed for children and teenagers than on tests constructed for adults. This would increase the number of different easy-level items and, therefore, increase the reliability of the discrimination as well as reduce transformation issues because the expected higher scores will be nearer the mean.

1.2.3. Different intelligence tests

Six intelligence tests for children and teenagers will be presented in this section. The tests were chosen for their popularity as they are often referred in scientific journals and books. Three of the tests are conventional intelligence tests which measure an array of cognitive skills using different subtests to assess verbal and nonverbal skills. These conventional tests were selected because they are widely used for scientific purposes and, therefore, have been thoroughly researched. The Wechsler Intelligence Scale for Children-Fourth Edition, the Stanford-Binet Intelligence Scale-Fifth Edition and the Kaufman Assessment Battery for Children-Second Edition were chosen as conventional intelligence tests.

Additionally, this review will present and analyse three nonverbal intelligence tests, each with a relatively wide age range compared to other nonverbal intelligence tests for children. Nonverbal intelligence tests are particularly suitable for testing individuals from different cultural backgrounds (McCallum, 2003). The presented nonverbal intelligence tests are: the Comprehensive Test of Nonverbal Intelligence, the Snijders-Oomen Non-verbal Intelligence Test 5½-17 and the Leiter International Performance Scale-Revised.

Each test will be briefly described and then analysed for its suitability to investigate the relationship between intellectual functioning and sports performance. The tests will be investigated for their validity, reliability, cultural bias and the duration of the assessment. As described in section 1.2.2., the decision to use tests designed for children and teenagers is expected to minimize issues concerning the use of intelligence tests for young adults with ID.

The Wechsler Intelligence Scale for Children

The first Wechsler Intelligence Scale for Children (WISC) was developed in 1949; the latest edition, the WISC-IV, was published in 2003. It is used for children from 6 years onwards up to the age of 16 years and 11 months. The WISC-IV contains four factors based index scores which are assessed with 15 subtests and additionally five supplemental subtests. The index scores are: Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index and Processing Speed Index (Flanagan & Harrison, 2005).

The subtests 'Vocabulary', 'Similarities', 'Comprehension', 'Information' and 'Word Reasoning' contribute to the Verbal Comprehension Index. This index measures verbal concept formation, verbal reasoning and comprehension. 'Vocabulary' consists of two different types of items, picture naming items and word definition items. For the subtest 'Similarities', the examinee will be presented with two words and has to name the common concept. 'Comprehension' examines the understanding of general principles and social situations. The subtest 'Information' measures general knowledge and in 'Word Reasoning' the examinee has to infer a common concept from a series of verbal clues (Wechsler, 2003).

The Perceptual Reasoning Index contains tests for fluid reasoning, spatial processing, attentiveness to detail and visual-motor integration. It includes the subtests 'Block Design', 'Matrix Reasoning', 'Picture Concepts' and 'Picture Completion'. For 'Block Design', the examinee has to copy a constructed model or picture using one- or two-coloured blocks. The subtest 'Matrix Reasoning' requires the examinee to complete a matrix of different figures by choosing the correct figure out of five response options. In 'Picture Concepts', the examinee looks at two or three rows of different objects and has to infer a concept, into which, one of the objects of each row fits. The subtest 'Picture Completion' assesses whether the examinee can find, which part of the picture is missing within a given time limit (Wechsler, 2003).

The Working Memory Index assesses mental capacity. Mental capacity refers to the ability of temporarily storing incoming information for calculation or transformation and to the ability to hold the outputs of these calculations or transformations. The Working Memory Index consists of the following three subtests: 'Letter-Number Sequencing', 'Digit Span' and 'Arithmetic'. For 'Letter-Number Sequencing', the examinee reads a sequence of letters and numbers and recalls the letters in alphabetical order and the numbers in ascending order. The subtest 'Digit Span' consists of two parts, repeating numbers forward and repeating them backwards after the examiner has read them out loud. For 'Arithmetic', the examinee has to orally solve arithmetic problems within a given time limit without the use of pen and paper (Wechsler, 2003).

The Processing Speed Index assesses how fast a person can process simple information without making mistakes. The Processing Speed Index is measured with the subtests 'Coding', 'Symbol Search' and 'Cancellation'. For 'Coding', the examinee looks at boxes containing pairs of shapes and numbers and then copies each number into a box with the shape that was paired with that number. In 'Symbol Search' the examinee looks at one or more target symbols and subsequently visually scans a row of symbols within a set time limit to search for the target symbol(s). 'Cancellation' requires the examinee to look at a random and a structured arrangement of pictures to identify a target picture within a given time limit (Wechsler, 2003).

The administration of the WISC-IV is based on a standard procedure. Some of the subtests have an age dependent starting point. The test is administered individually, and the assessment takes about 60 to 90 minutes.

The standardization research of the WISC-IV was based on a sample of 2200 children, with 200 in each age group (The Psychological Cooperation, 2003). The average coefficients for internal consistency reliability varied between .72 to the .94 depending on age group. The mean coefficients for test-retest reliability varied between .86 and .93. Construct validity was

assessed using factor analysis (Keith, Goldenring Fine, Taub, Reynolds & Kranzler, 2006), which showed that the WISC-IV measured the same construct (four factor model) for all ages. There are limited data available concerning the criterion validity. The correlation coefficient between the full scale IQ WISC-IV scores and the Wechsler Intelligence Scale for Children-third edition (WISC-III) is .89 (Flanagan & Kaufman, 2004). Although the WISC-IV was thoroughly reviewed and adapted during its development in order to minimize cultural bias (The Psychological Cooperation, 2003), the bias analysis was limited to different cultures within the USA. While this might minimize any bias concerning different cultural backgrounds and languages spoken in the USA, it would not reduce the bias for non-American populations. Therefore, it cannot be assumed that the cultural bias has been minimized for the use of the WISC-IV across different countries worldwide.

Considering the strong emphasis on verbal ability and general knowledge (Verbal Comprehension Index) and the relative long duration of the assessment (60 to 90 minutes), the WISC-IV is an unlikely candidate for the research project. Furthermore, although the developers of the WISC-IV tried to minimize the cultural bias of the test, this has only been done for different cultures within the USA and not on a wider, global level. Taking the above considerations into account, the WISC-IV was not an option for this project.

The Stanford-Binet Intelligence Scale: Fifth Edition

The fifth edition of the Stanford-Binet Intelligence Scale was published in 2003 (Roid, 2003a). The test was originally developed by Alfred Binet and Theophile Simon in 1905. Since it was revised under the direction of Lewis M. Terman 1916 at Stanford University/USA, it is called the Stanford-Binet Intelligence Scale (Cohen & Spenciner, 2003). From its very early days onwards, it was developed as a scale that could measure participants from a

36

wide age range and the current edition was designed for an age range between 2 and 85+ years.

The Stanford-Binet Intelligence Scale Fifth Edition (SB5) test battery measures intellectual ability. Based on the Cattell-Horn-Carroll theory, it assumes a hierarchical general factor model with the factor "g" contributing to five different cognitive factors (Roid & Pomplun, 2005). These five factors are: 'Fluid Reasoning', 'Knowledge', 'Quantitative Reasoning', 'Visual-Spatial Processing' and 'Working Memory'. Each factor is measured with a set of verbal as well as non-verbal subtests.

'Fluid Reasoning' is the ability to solve new problems. It is assessed with the verbal subtests 'Early Reasoning', 'Verbal Absurdities' and 'Verbal Analogies' and the non-verbal subtest 'Object-Series/Matrices'. 'Early Reasoning' is a test for which the examinees have to identify cause-and-effect relationships in pictures. In 'Verbal Absurdities' the ability to verbally express absurd contradictions is assessed. In the test 'Verbal Analogies', the examinees have to find an underlying concept of object pairs. The non-verbal subtest 'Object-Series/Matrices' is a routing subtest, which is administered in the beginning of the test series to determine at which level the examinee starts.

'Knowledge' refers to the factor that describes the collection of general knowledge of a person. It consists of one verbal and two non-verbal subtests. The verbal subtest ('Vocabulary') is again a routing test to set the starting level of the examinee. The first non-verbal subtest is 'Procedural Knowledge', for which the examinee has to show that he knows the presented objects by using gestures to describe it. In the second subtest, 'Picture Absurdities', the examinee is asked to point to pictures and to explain what is unusual about them.

'Quantitative Reasoning' refers to the ability to solve numerical and word problems. It consists of 'Verbal Quantitative Reasoning' and 'Non-verbal quantitative Reasoning'. In 'Verbal Quantitative Reasoning', the examinee is

37

presented with numerical concept and word problems. 'Non-verbal Quantitative Reasoning' measures the ability to solve numeric problems.

'Visual-Spatial Processing' refers to the capacity to see associations between objects, describe spatial orientation and recognize patterns in visual items. The verbal subtest 'Position and Direction' measures the understanding and ability to follow directions. The examinee is asked to follow given spatial directions on a map. The first non-verbal subtest 'Form Board' is a simple structured puzzle, the second non-verbal subtest is 'Form Patterns', in which the examinee is asked to use pieces to shape people, animals or objects.

'Working Memory' is the ability to keep visual as well as verbal information in the short-term memory and then convert or transfer it. It consists of two verbal and two non-verbal subtests. In the verbal subtest 'Memory for Sentences', the examinee has to attempt to remember all words in a sentence. For the other verbal subtest, 'Last Word', the examinee has to remember the last word in a series of sentences. The non-verbal subtests are 'Delayed Response' and 'Block Span'. 'Delayed Response' measures the ability to remember which toy is hidden under different plastic cups, and in 'Block Span' the examinee is asked to tap different blocks in a certain sequence with the number of blocks increasing to increase the difficulty of the task.

Although the SB5 contains 16 subtests, the majority of participants will not be required to complete all the tests. The administration of the test battery is described in three consecutive books (item book 1 to 3). Item book 1 starts with the two routing subtests ('Object-Series/Matrices' followed by 'Vocabulary'), which set the level for the examinee which will determine where to start in book 2 (non-verbal subtests) and book 3 (verbal subtests). Unlike the Wechsler Intelligence Scales, the assessment of the SB5 is not a set procedure. The examiner is explicitly required to consider the examinee's background, especially their linguistic background. Based on that

consideration, the examiner might decide for a non-verbal administration of the intelligence test, leaving out the verbal subtests. The assessment with the nonverbal SB5 is shorter (around 40 minutes) and shows consistently high correlation coefficients (between .94 and .97, depending on age group) with the full IQ scale. The assessment of the standard SB5 takes between one and two hours and is individually administered.

The normative sample for the standardization of the SB5 included 4800 individuals with an age range of 2 to 96 years. The research report of the SB5 (Roid, 2003b) indicated an internal- consistency reliability ranging from .95 to .98 for IQ scores and from .90 to .92 for the five factor index scores. Internal-consistency reliability for the subtests ranged from .84 to .89. Test-retest reliability ranged depending on age group from .93 to .95 for IQ scores.

The validity of the SB5 is established quite well according to the research report (Roid, 2003b). Studies show a high correlation between the SB5 and other achievement tests, such as the Wechsler Individual Achievement Test. Factor analyses supported the five factor model, which is the theoretical foundation of the test (Roid, 2003b).

Although the SB5 was developed, analysed and reviewed in order to reduce culture bias, those studied were different cultures within the USA only (Roid & Pomplun, 2005). Therefore, a cultural bias for non-Americans cannot be excluded.

The standard SB5 intelligence test has a strong emphasis on language and general knowledge. Furthermore, the duration of the assessment (one to two hours) is too long for the purpose of this research project. Although the nonverbal version is considerably shorter (40 minutes) and uses nonverbal subtests only, it still tests knowledge with the subtests 'Procedural Knowledge' and 'Picture Absurdities'. The knowledge tested in these subtests might be culture dependent. Further research will have to establish to what extent cultural bias influences the outcomes when the test is used for an

international population. Considering the duration of assessment and the issue of cultural bias, this test was not deemed appropriate for the project.

The Kaufman Assessment Battery for Children: Second Edition

The second edition of the Kaufman Assessment Battery for Children (KABC-II) is the successor of the Kaufman Assessment Battery for Children which was first published in 1983 (Kaufman & Kaufman). The KABC-II assesses information processing and cognitive abilities of children between 3 years and 18 years and 11 months. It is based on the theoretical model of Luria, which assumes three functional units and the Cattell-Horn-Carroll theory, which categorizes specific cognitive abilities into different factors (Kaufman, Kaufman, Kaufman-Singer & Kaufman, 2005). The KABC-II consists of core and expanded test batteries, using the expanded batteries to widen the range of the cognitive abilities assessed with the core battery. Depending on the examinee, it has to be decided prior to the assessment whether the KABC-II is administered according to Luria's Mental Processing Index (MPI), or according to the Fluid-Crystallized Index (FCI), based on the Cattell-Horn-Carroll model. The latter is the option applied to the majority of examinees. It is also used when the examinee has learning or ID, has emotional or behavioural disorders, is diagnosed with attention-deficit/hyperactivity disorder or is assessed for giftedness. On the other hand, the MPI is the preferred option if the child is from a multi-lingual or non-mainstream cultural background, has language disorders, is diagnosed or suspected to be autistic or if for any other reason acquired knowledge should not be influencing the test results. The choice of index (FCI or MPI) and age of the child will determine the selection of subtests. The KABC-II offers 17 different verbal and non-verbal subtests, which are categorized into five sub-indices.

The 'Sequential Index' measures short-term memory and consists of three different subtests: For 'Word Order' the examinee is required to touch a

series of silhouettes of common objects when they are called out by the examiner. In 'Number Recall' the examinee repeats a series of numbers. 'Hand movements' require the examinee to repeat the exact number of taps on the table copying the examiner.

The 'Simultaneous Index' assesses visual processing with six different subtests. For the subtest 'Triangles', the examinee matches an abstract picture with coloured triangles. The subtest 'Face Recognition' requires the examinee to look at a picture with one or two faces and then to choose a picture with the same face(s) from a set of photographs. 'Conceptual Thinking' assesses whether the examinee can identify the one picture that does not belong into a set of pictures. In 'Rover', the examinee has to move a toy dog on a checkerboard-like game towards a bone, avoiding obstacles and choosing the path with the fewest moves. 'Block Counting' tests the ability to count blocks in pictures where these are partially hidden. For 'Gestalt Closure', the examinee has to mentally fill in missing parts of an inkblot drawing and then name the perceived picture.

The 'Planning Index' is designed to measure the ability to plan and to program behaviour. It is assessed with two subtests, 'Pattern Reasoning' and 'Story Completion'. For 'Pattern Reasoning', the examinee is presented with a series of stimuli in order to form a logical, linear pattern. However, to complete this pattern one more stimulus is needed. At the bottom of the page are several options (between four and six). The examinee has to choose the option that completes the pattern correctly. The subtest 'Story completion' tests if the examinee can complete a story consisting of a series of pictures. Some pictures are missing and have to be chosen from a set of options.

The 'Learning Index' examines the capacity of long-term storage and retrieval with the subtests 'Atlantis', 'Rebus' and 'Delayed Recall'. In 'Atlantis', the examinee is taught nonsense names for sea life creatures, which have to be recalled when the examiner later points to them. For 'Rebus', the examiner presents the examinee with different drawings (rebus) and names words or concepts for these. Later the examinee has to remember these words or

concepts in order to be able to build a sentence out of the words for the different drawings. 'Delayed Recall' tests how many paired associations learned 20 minutes earlier during the 'Atlantis' and 'Rebus' subtests an examinee can recall.

The 'Knowledge Index' measures general knowledge using the subtests 'Riddles', 'Expressive Vocabulary' and 'Verbal Knowledge'. For 'Riddles', the examinee is presented with clues of a concrete or abstract concept, which then have to be pointed out. 'Expressive Vocabulary' assesses if the examinee can provide the name of a pictured object. For 'Verbal Knowledge', the examinee has to choose one item out of a set of six pictures, which relates to a presented word, or to answer a general knowledge question.

As explained above, the examiner will firstly choose the model on which the subsequent test should be based. If necessary, the examiner can also choose to use a non-verbal version of the KABC-II, which is a combination of all non-verbal subtests. Their instructions are then given in pantomime. Depending on the age of the examinee and the index used (FCI or MPI), the assessment takes between 30 and 70 minutes. The tests are administered individually.

All psychometric properties were assessed by Kaufman and Kaufman (2004). The standardization sample for the KABC-II included 3025 children and adolescents with 125 to 250 individuals for each year. Internal consistency reliability was assessed using split-half reliability coefficients. For the MPI, the mean reliability coefficient for the age groups 3-6 and 7-18 was .95, for the FCI this was .96 and .97, respectively. The non-verbal scale showed a mean internal consistency coefficient of .90 for 3 to 6 year-olds and .92 for 7 to 18 year-olds. Test-retest reliability coefficient for an interval period of one month was depending on the age group for the MPI between .86 and .91, and, for the FCI, between .90 and .94.

The results of a factor analysis study support the construct validity (Klanderman, Devine & Mollner, 2006). Correlation coefficients between the

FCI and the Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV) were .89 for 7 to 16 year-olds, between the FCI and the Woodcock-Johnson III Tests of General Intellectual Ability .72 for ages 7 to 16. Correlations of the MPI with these intelligence tests were about .05 points lower.

Similar to the other tests discussed, an ethnicity analyses was limited to different minority cultures within the USA only (Kaufman, 2003).

Like in the WISC-IV and the SB5 intelligence tests, language and general knowledge play an important role in the KABC-II test. Although the KABC-II offers a nonverbal version using pantomime for communication between examiner and examinee, the influence of its cultural bias has not yet been satisfactorily established. Additionally, none of the competing indices are ideal for this project as the MPI is the preferred option for individuals from different cultural backgrounds and the FCI is recommended for individuals with ID. The duration of the assessment is shorter than for the WISC-IV or the SB5. However, depending on cognitive ability, the assessment might still take more than one hour per participant. For these combined reasons, the KABC-II was not be considered for this project.

The Comprehensive Test of Nonverbal Intelligence

The Comprehensive Test of Nonverbal Intelligence (CTONI) was first published in 1997 (Hammill, Pearson & Wiederholt). The CTONI measures nonverbal abstract problem solving and reasoning for an age range between 6 years and 18 years and 11 months. The test is not built on a theoretical foundation, but is derived from an analysis of the items of other nonverbal tests. This analysis resulted in several principles concerning the test instructions, the abilities the test should measure, and the type of items the test should include (Braden & Athanasiou, 2005).

The test consists of six different subtests: 'Pictorial Analogies', 'Geometric Analogies', 'Pictorial Categories', 'Geometric Categories', 'Pictorial Sequences' and 'Geometric Sequences' (Pearson, 2003).

'Pictorial Analogies' and 'Geometric Analogies' measure the ability to recognize the relationship between two objects. The examinee is presented with a set of two pictures or geometrical figures. Under this set is another picture/geometrical figure and an empty frame. Then the examinee is asked to choose one of five pictures/geometrical figures for the empty frame, so that the second set equals the relationship of the first set.

'Pictorial Categories' and 'Geometrical Categories' assess the ability to infer resemblance. The examinee is presented with two different pictures/geometrical figures, which have something in common and one empty frame. The examinee then has to choose from five pictures/geometrical figures the one that is most similar to the first two.

'Pictorial Sequences' and 'Geometrical Sequences' measure the ability to logically complete a sequence. The examinee is shown a sequence of three pictures/geometrical figures and one empty frame. The examinee then has to choose one out of five pictures/geometrical figures to complete the sequence plausibly.

The CTONI can be administered as a 'paper' version or as a computerized version. The instructions for all items can either be pantomimed or be given verbally by the examiner. Responses can be given by pointing ('paper' version) or by clicking the computer mouse (computer version). All examinees start with the first item of each subtest and continue until they have reached a certain number of incorrect items. There are no different starting points for different age groups. The test is based on a standard procedure, which is identical for all examinees. The assessment takes between 20 and 45 minutes. Each examinee is tested individually.

The normative standardization sample for the CTONI involved 2901 individuals ranging from 6 years to 18 year and 11 months (Hammill & al,

1997). The internal consistency coefficient varied for the total test between .95 and .97 and, for the subtests, between .85 and .94, depending on age group (Braden & Athanasiou, 2005). Test-retest reliability coefficients varied, again depending on age group, between .79 and .94 (Cohen & Spenciner, 2003).

Criterion validity was established by comparing the IQ scores of the CTONI with other intelligence tests (Pearson, 2003). The correlation coefficient with the WISC-III for learning disabled children was .81 and for deaf children .90. Correlations with the Wechsler Scales for non-disabled children were not provided. The correlation coefficient between the CTONI and other nonverbal intelligence tests (not specified) for the general population was .80. Construct validity was supported, as the results of a factor analysis showed that all subtests loaded on a single factor representing the general intelligence factor "g" (Pearson, 2003).

Cultural bias was assessed by comparing the results of different minority groups in the USA again (Pearson, 2003). The mean IQ scores ranged from 95 for Indian Americans to 103 for Asian Americans. Comparisons with different cultural groups outside the USA have not been published.

The duration of the administration is rather short (20 to 45 minutes), which would fit very well with the selection criteria. However, there are no data investigating the results for cultures outside the USA. Therefore, the degree of cultural bias has not been sufficiently established. Nevertheless, this test will be considered in the conclusions because of its short duration.

The Snijders-Oomen non-verbal intelligence test 5½ - 17

The first Snijders-Oomen non-verbal intelligence test was published in 1943 and was designed exclusively for deaf children. The current version, the

Snijders-Oomen non-verbal intelligence test 5½ - 17 (SON-R) has been published in 1989 and is a standardized version for both hearing and deaf children from five and a half to seventeen years old.

The SON-R is not based on a theoretical construct or model, but developed on the base of a combination of empirical and theoretical considerations concerning the intellectual abilities a nonverbal intelligence test should and could measure (Snijders, Tellegen & Laros, 1989). Consequently, the SON-R contains four different types of subtests: for abstract, concrete, spatial and perceptual reasoning.

The tests measuring abstract reasoning are 'Categories' and 'Analogies'. For the subtest 'Categories', the examinee is presented with three pictures of objects that are related to each other. The examinee is then asked to choose from a set of five more pictures those two objects that share common features or characteristics with the first three. 'Analogies' requires the examinee to discover the principle behind the change of a geometrical figure and to apply that change to another geometrical figure using one of four possible options.

Concrete reasoning abilities are assessed with the subtests 'Situations' and 'Stories'. For the subtest 'Situations', the examinee is presented with a drawing of which one or more parts are missing. The examinee then has to choose one or more parts from a number of alternatives in order to complete the picture in a coherent way. 'Stories' assesses if the examinee can find the right sequence to a set of four cards in order to form a story.

The spatial subtests are 'Mosaics' and 'Patterns'. For the subtest 'Mosaics', the examinee is asked to copy various mosaic patterns from a picture into a two-dimensional frame using red, white and patterned squares. In the subtest 'Patterns', the examinee is shown a continuing line, which has a systematic pattern. One part of the line is missing. The examinee is required to fill in the missing part in accordance with the rest of the line.

Perceptual abilities are measured with the subtest 'Hidden Pictures'. For 'Hidden Pictures' the examinee is shown a simple picture, which is also

46

hidden several times in a drawing. The examinee is asked to point out the hidden pictures in the drawing. There are four different search drawings.

The instructions to all subtests can be given verbally or pantomimed. All examinees start with the same item in each subtest, regardless of their age. All subtests, with exception of 'Hidden Pictures', consist of two or three series of items. Each series starts with an easy item and gets increasingly more difficult. In each subtest the starting point for the consecutive series depends on the score of the previous series. Only for the subtest 'Hidden Pictures' do all four search drawings have to be completed.

The SON-R can also be administered in a shortened version with only the four subtests 'Categories', 'Analogies', 'Situations' and 'Mosaics'. The advantage of using the shortened version is the reduced administration time. The disadvantages are a lower mean reliability of the total score (.90 compared to .93 for the complete version) and a lower mean generalisability to the domain of comparable subtests (.77 compared to .85 for the complete version) (Snijders & al., 1989).

The duration of the complete version of the SON-R is on average 79 minutes, the duration of the shortened version is 38 minutes on average. The test is administered individually.

The standardization sample included 1350 individuals between 6 and 14 years old living in the Netherlands, with each age group containing 150 individuals (Snijders & al., 1989). The age range was widened by extrapolation to 5 ½ to 17 years of age. The reliability coefficient of the total score varies from .90 to .94, depending on age group and the generalisability coefficient from .79 to .89 depending on age group (Tellegen & Laros, 1993). Research concerning the test-retest reliability has not been published to our knowledge.

As mentioned above, the reliability coefficients for the shortened version of the SON-R are lower. The reliability coefficient for the total score varied between .85 and .91, depending on age group and the generalisability

47

coefficient varied between .67 and .82 (Tellegen & Laros, 1993). Again, there are no publications concerning test-retest reliability.

Validity was assessed by relating scores on the SON-R to school achievement and other general intelligence tests, such as the Wechsler Intelligence Scale for Children. The correlation between scores on the SON-R and school teacher's evaluation was only .33. The correlations between the SON-R score and school report, as well as a Dutch primary school assessment test (CITO) for 11 year olds, were both times .66 (Snijders & al., 1989). A study with 35 children comparing outcomes of the SON-R, the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the Ravens Progressive Matrices Test (Nieuwenhuys, 1991) showed no significant differences between the means of IQ scores.

Cultural bias was assessed by comparing the mean IQ scores of children from parents who were not born in the Netherlands. Mean IQ scores ranged from 101.4 for the African-Asian-American parental background, to 82.7 for children of Turkish parents. When these data were controlled for parent's occupation, the differences decreased substantially (Snijders & al., 1989). Additionally, cultural bias was evaluated in a comparison of Brazilian and Dutch children (Tellegen & Laros, 2004). The children of both groups were assessed with those tests containing pictorial representations as these could be culture dependent. When an item was scored incorrectly the child was asked whether they had recognized the item. The results showed that 8 of the 80 items (one item in 'Categories', four items in 'Situations' and three items in 'Stories') were culturally biased.

The administration time of the shortened version of the SON-R is quite short (38 minutes on average), which would be suitable for the application procedure. Cultural bias has not only been studied by comparing the scores of different minority groups within the country of the origin of the test but also through an investigation of items and scores in two different countries. This analysis also had identified five culturally biased items that are included in the

48

shortened version of the SON-R. These items would either have to be changed or deleted from the test in order to eliminate the cultural bias of the test for the athletes. Although this would make it necessary to re-norm the test, it would also provide this test with a very advanced ethnicity analysis compared to the other intelligence tests.

Although the reliability for the shortened version of the SON-R is lower than for the complete version, the overall reliability is still similar to the mean reliability of the other reviewed nonverbal intelligence tests. Considering the short duration of the assessment time and the fact that culturally biased items have already been identified, this test will be taken into account in the conclusions.

The Leiter International Performance Scale-Revised

The Leiter International Performance Scale-Revised was developed on the basis of the original Leiter International Performance Scale with the intention to design an intelligence test for IQ assessment in different cultures (Roid, Nellis & McLellan, 2003). It measures intelligence for an age range from 2 years to 20 years. The Leiter-R is based on the hierarchical models of Carroll (1993) and Gustafsson (1988). Although the Leiter-R offers 20 subtests divided into the two categories 'Visualization and Reasoning' (VR) and 'Attention and Memory' (AM), only the subtests in the category VR are used to obtain the IQ score. The AM subtests provide a separate measure to assess attention and memory abilities. The subtests used to assess the IQ score are: 'Sequential Order', 'Repeated Patterns', 'Figure Ground', 'Form Completion', 'Matching', 'Classification', 'Design Analogies' and 'Paper Folding' .

For the subtest 'Sequential Order', the examinee has to choose a picture or figure in order to complete a sequence logically. The subtest 'Repeated Patterns' shows a pattern with a missing part, which the examinee

49

has to complete with a card chosen from a number of possibilities. 'Figure Ground' is a subtest that requires the examinee to find a number of hidden figures within a complex picture. For 'Form Completion', the examinee has to recognize an object when presented with fragments of it. 'Matching' requires the examinee to select the correct response card, in order to match the stimulus. In the subtest 'Classification', the examinee is expected to categorize presented objects according to a common concept. For 'Design Analogies', the examinee is presented with a matrix of geometrical shapes which the examinee has to complete in a logical way using one of the possible options. 'Paper Folding' is the mentally folding of a two-dimensional shape into a target shape.

The Leiter-R can be administered in a short form which uses the subtests 'Sequential Order', 'Repeated Patterns', 'Figure Ground' and 'Form Completion'. This abbreviated form will take about 25 minutes to administer. The full scale Leiter-R has a duration of about 40 minutes and includes six subtests. To achieve a full scale IQ for children aged 2-5, the subtests 'Sequential Order', 'Repeated Patterns', 'Figure Ground', 'Form Completion', 'Matching' and 'Classification' are administered. For the age group 5-20 year olds, 'Sequential Order', 'Repeated Patterns', 'Figure Ground', 'Form Completion', 'Design Analogies' and 'Paper Folding' are tested.

The Leiter-R is administered individually. The instructions of the subtests are pantomimed. Each subtest starts with initial teaching items in order to clarify the instructions for the examinee.

All psychometric properties were evaluated by Roid & Miller (1997; 1999). Internal consistency reliability coefficients ranged for six subtests of the full IQ scale from .91 to .93 and from .88 to .90 for the shortened version. For the VR subtests, the internal consistency coefficients for the different age groups ranged from .75 to .90. Test-retest reliability coefficients were between .90 and .96 for the full scale version and between .88 and .96 for the shortened version. Content validity was verified by independent reviewers.

Criterion validity was evaluated by comparing the scores of the Leiter-R full scale IQ results to the WISC-III results. The correlation coefficients were .86 between the Leiter-R full scale IQ results and the WISC-III, and .85 between the short version of the Leiter-R and the WISC-III. A factor analysis of the subtests showed that the number of factors changed depending on age group. While there were only three factors for the age group 2-3 year olds, the number increased to six factors for the age group 6-10 year olds, and was reduced to five again for the age group 11-20. This factor model compared to other established factor models (Roid & Woodcock, 2000). Although the Leiter-R was developed to be a suitable measure for IQ assessment in different cultures, ethnicity studies were again limited to different cultures within the USA (Roid & al. 2003). The mean IQ score of the brief scale ranged from 98.0 for Navajo children in Arizona, to 102 for Hispanic Americans. A cultural bias for non-American cultures cannot be excluded.

The duration of the test (25 minutes) is convenient for the purpose of the project. However, there are no studies examining the Leiter-R for cultures outside the USA. Therefore the degree of cultural bias has not been sufficiently established. However, due to the short duration of the assessment, the test will be considered in the conclusions.

1.2.4. Interim conclusion

The application procedure for intellectually disabled athletes for the Paralympics requires an intelligence test that is suitable for both teenagers and young adults with ID. A short duration of test administration is preferable as it facilitates the use during sport events. As the Paralympic athletes come from many different countries and cultures, it is important that the test has a minimal cultural bias.

Until now, there is no intelligence test designed for teenagers and young adults with ID. In order to provide an intelligence test that offers sufficient numbers of different level easy items and, therefore, can avoid 'floor' effects, it was decided to use an intelligence test developed for the assessment of children and teenagers. The review of three conventional and three nonverbal intelligence tests revealed that nonverbal tests take considerably less time to administer than conventional intelligence tests. The Leiter-R (brief form) is on average the quickest test (25 minutes), followed by the CTONI (20-40 minutes) and the shortened version of the SON-R (38 minutes). An additional advantage of the CTONI is the existence of a computerized version of the test, which would make it particularly easy to use for large scale assessment.

Another advantage of the nonverbal intelligence tests is that they do not test general or verbal knowledge, which is culturally dependent. Therefore, they are culturally more insensitive than conventional intelligence tests (Braden & Athanasiou, 2005; Lopez, 1997; Seguin, 1907).

Although reliability and validity coefficients were given for all tests (except the test-retest reliability for the SON-R), the results cannot be transferred to a different population and, therefore, have to be re-evaluated for athletes with ID.

For all three nonverbal tests, the cultural bias has been assessed. The biggest difference between different cultures in mean IQ score was found in the SON-R (18.7 IQ points), while the difference between cultures for the Leiter-R intelligence test was only 4 IQ points. However, it is important to note that only the cultural bias of the SON-R was evaluated in different countries, while the degree of cultural bias for the CTONI and Leiter-R was assessed for different cultures within the USA only. Therefore, conclusions concerning their degree of cultural bias have to be drawn with caution. Only for the SON-R, cultural fairness has been evaluated with children outside the country it has been standardized in, and culturally biased items have been identified. This gives this test the advantage to ensure its cross-cultural insensitivity.

In summary, all three nonverbal intelligence tests are suitable for further research to develop an application procedure. The duration of the shortened version of SON-R is longer than the other nonverbal tests, but the cultural bias of its items has already been identified and, therefore, does not need to be examined again. The CTONI has the advantage of a computerized version and the Leiter-R the shortest administration time, but both tests need to be assessed for their cultural bias in countries outside the USA. Consequently, all three tests are suitable for research to develop a classification system for the Paralympics.

1.3. Discussion

Research studies indicated that top-level athletes with ID perform less well on physical fitness tests than physically trained individuals without disabilities. These results suggested that intellectual disabilities are associated with physical performance. However, most of these studies controlled only for some possibly confounding factors and not for motivation and lifestyle differences.

The studies comparing elite athletes with and without ID assumed to control for lifestyle factors, motivation and differences in training, as athletes, in particular elite athletes, are thought to have a healthy lifestyle, to be ambitious and to have professional training. These assumptions, however, have never been verified. Intellectually disabled athletes will probably have a healthier lifestyle and might be more motivated and trained than non-athletes. But they still might have a very different lifestyle and might be less motivated and trained compared to non-disabled athletes. Therefore, it is unclear if physical performance studies of elite athletes with ID need to be controlled for lifestyle factors, motivation and training received.

Research into underlying reasons for the association between intellectual disability and physical performance indicated that there are several possible biological causes. Some of these are genetic, like Down

53

syndrome, others are developmental, such as impairment of the cerebellum. All of these affect the intellectual, as well as the physical development of the individual. Considering that there are several possible reasons for the relationship between intellectual disability and physical performance, it is likely that the relationship will not the same for all individuals with intellectual disability. Depending on the cause for the disability, it might mean that some athletes with ID are physically much more capable than others, even if they perform identically on an intelligence test. Further research with separate groups for the different causes of intellectual disability could investigate if the cause of the disability (as far as it actually can be established) affects the relationship between cognitive impairment and physical performance.

The study of conventional and nonverbal intelligence tests showed, firstly, that there are large differences in administration time between the tests. The nonverbal intelligence tests had a considerably shorter administration time than conventional tests. Although a short administration time avoids issues with attention problems of the target population and is convenient for the assessment of large numbers of participants at an international sports event, it raises concerns about the capacity of the test to differentiate between the different levels of cognitive functioning. A short intelligence test has a limited number of items. Consequently, the transformation from raw score into standardized IQ scores will only allow a rough estimation, but not a good discrimination between IQ levels. Therefore, the discriminative capacities of a very short intelligence test will have to be further investigated.

Conventional intelligence tests are likely to be more culturally biased than the nonverbal tests, as they require general and verbal knowledge which depend on school curricula and home education. Both are largely influenced by the culture and values of a country. Many nonverbal intelligence tests praise themselves to be cross-culturally insensitive, but their cultural bias has often only been researched for the population of the country in which the test has been developed. Particularly, items using pictures can easily be culturally

biased as the study by Tellegen and Laros (2004) demonstrated. This might not always be noticed by investigating different cultures living in one country, because they still share common influences. Additional investigations regarding the cultural bias will be necessary if any of the tests are to be used in the application procedure for the Paralympics.

Nevertheless, any of the presented nonverbal tests would be suited to explore the relationship between intellectual functioning and physical performance. The SON-R intelligence test was chosen, as a cultural fairness study had already identified a limited number of culturally biased items, which can be taken into account in the following studies.

Chapter 2: The impact of intellectual impairment on physical and sport performance

2.1. Methods for pilot project and studies 1 to 5

The following studies investigated, for elite athletes with ID, the relationship between the degree of intellectual functioning and the level of physical performance. An initial pilot study focused on recreational football players with ID. Studies 1 to 3, however, assessed athletes with ID competing in sports disciplines that are currently being considered by the IPC for re-inclusion in the category 'Intellectual Disability': track and field athletics; table tennis; and swimming.

As discussed in chapter 1, several studies suggested that depending on sports discipline, different cognitive abilities predict physical performance of non-disabled athletes (Kasahara, Mashiko & Niwa, 2008; Ryan, Atkinson & Dunham, 2004; Overney, Blanke & Herzog, 2008). Therefore, the above relationship was investigated using both the results for individual SON-R subtests measuring different cognitive abilities, as well as the IQ score derived from the full test results. In addition, the results were analysed for each different sports discipline and also for the overall sample of athletes.

Based on previous research the alternative hypothesis is: *There is a positive association between IQ scores and physical performance for elite athletes with ID.*

Studies 4 and 5 evaluated reliability and validity of the SON-R intelligence test for individuals with ID. The SON-R research report investigated reliability and validity for a population without ID (Snijders & al., 1989) and, therefore, it is necessary to re-evaluate these qualities of the test for individuals with ID (Jensen, 1980).

Reliability is the degree to which a test achieves repeatability of values or scores (Bartram, 1990). Reliability for a test is not expressed in a single value but is based on a set of reliability studies, which together will provide an estimation of the reliability of an instrument (Kline, 2005). Reliability for the SON-R was evaluated for individuals with ID. The evaluation of reliability in this study included test-retest reliability and inter-rater reliability.

Test-retest reliability concerns the stability of a test over time and is computed using a correlation analysis. There are several concerns regarding the test-retest reliability, the most obvious being the training effect of a repeated test administration. Although the training effect of cognitive ability tests is established, its effect size for the mere repetition of the test is relatively small (Cohen's d =.26; Hausknecht, Halpert, Di Paolo & Moriarty Gerrard, 2007). Another problem when assessing test-retest reliability is the lack of control for internal consistency. Depending on the structure of the test high test-retest reliability does not exclude a very different pattern of scoring on an individual basis for the two test administrations (Nunnally & Bernstein, 1994). For example, a score of 14 correct answers could be reached in the retest session by scoring 14 completely different items correctly. Therefore, it is important that internal consistency reliability is established for a test as well as test-retest reliability. However, for the evaluation of internal consistency each item of the parallel versions of the SON-R subtests would have to be administered. This was not possible due to practical restrictions. Therefore, internal consistency was not evaluated for the use with individuals with intellectual disabilities. This study assessed test-retest reliability of the SON-R for individuals with ID.

The alternative hypothesis was that *test-retest reliability can be demonstrated for the administration of the SON-R for individuals with ID.*

Inter-rater reliability refers to the influence of the person who administers the test on the test result. The scoring system should prevent any

influences by the test assistant, but the possibility of such influences cannot be excluded without further investigation. Inter-rater reliability is examined with two test assistants administering the test to the same participants. The correlation between the scores of the two administrations will then give an estimate of the inter-rater reliability. Although the SON-R provides an objective, standardized scoring system, the test assistant could influence the answer through nonverbal clues or affect the performance in other ways. Consequently, the influence of the test assistant on the test outcome should be assessed. This study analysed inter-rater reliability of the SON-R for individuals with ID.

The alternative hypothesis was that *inter-rater reliability can be confirmed for the use of the SON-R for individuals with ID.*

The validity of an instrument determines to what extent the instrument measures what it is supposed to measure (American Educational Research Association, 1999). For psychological tests, validity cannot be expressed in a single value, but needs the accumulation of evidence supporting validity, which will indicate a degree of validity rather than an 'all-or-nothing' property (Aiken, 1994; Nunnally & Bernstein, 1994). Furthermore, validity research will not evaluate the instrument itself but the use of the instrument in a certain context or for a specific purpose (Nunnally & Bernstein, 1994). There are several methods of validity that will be investigated to evaluate the overall validity of the SON-R:

Content validity is a theoretical consideration that refers to the extent to which the items of an instrument represent the concept which the instrument measures. The content validity of the SON-R has been confirmed in its research report (Snijders, Tellegen & Laros, 1989) and is unlikely to change for the use for individuals with ID.

Construct validity refers to the relationships between scores within a test. These relationships are expected to represent the underlying concept of the test (Cronbach, 1990). The concept has to be precisely defined in order to

be examined for its construct validity. The definition of the construct will be the basis for one or more hypotheses which can then be tested. Depending on the complexity of the construct, various methods can be employed to examine its validity: For example, outcomes of two or more different groups whose scores are expected to differ can be compared. Alternatively, a factor analysis can determine if a test does indeed reflect the number of ability factors the construct suggested. The construct validity of the SON-R was assessed with a principal component analysis. For a non-disabled population, the SON-R showed one dominant factor (Snijders, Tellegen & Laros, 1989). Therefore, the alternative hypothesis was that *construct validity will be confirmed for the use of SON-R for the assessment of individuals with ID.*

Criterion validity refers to the extent to which a test predicts or correlates to a certain criterion. Criteria can be academic success, work related performance measures or outcomes of follow-up studies (Cronbach 1990; Kline, 1993). This study investigated criterion validity for individuals with ID based on a correlational analysis with existing IQ scores. Criterion validity of the SON-R will be evaluated in a comparison with registration IQ scores (scores on the Wechsler Adult Intelligence Scale or the Wechsler Intelligence Scale for Children, depending on age of the participant). As both tests are designed to measure intelligence, the alternative hypothesis was that *there is an association between scores of the SON-R and scores on Wechsler Scale for individuals with ID.*

2.1.1. Participants

Participants included in the pilot project attended a special needs department of a local college and took part in an inclusive football program. In total, 16 young adults between 17 and 23 years took part in this study. Four participants were female, twelve were male. One male participant had tunnel vision and was excluded from the analyses, leaving 15 participants in total.

All participants included in studies 1 to 5 were athletes training for the England squad in the category 'Intellectual Disability', or aspirants who were screened for their eligibility for the category 'Intellectual Disability'. All athletes were invited to national training camps or sports events by MENCAP charity[1] and included individuals with ID as well as other learning disabilities, such as autism or attention deficit hyperactivity disorder. Therefore, the sample included athletes with IQ scores above the threshold of 75.

Table 1 summarizes the gender distribution and age of the participants for the pilot study and studies 1 to 3.

Table 1 Distribution for gender and age for pilot project and study 1 to 3

	Male	Female	Age	N
Pilot project	11	4	17-23	15
Track and field athletics	14	3	14-40	17
Table tennis	16	8	15-50	24
Swimming	13	5	8-40	18

The evaluation of reliability of the SON-R for individuals with ID (study 4) included 14 table tennis players (10 male, 4 female) of study 2 who were retested on two different occasions to establish test-retest and inter-rater reliability of the SON-R for individuals with ID. Their age ranged between 17 and 50 years.

The validity of the SON-R for individuals with ID was assessed based on construct and criterion validity (study 5). Construct validity was investigated using the scores on the different subtests of 91 participants (63 male, 28 female) between 8 and 50 years of age. These participants included the 74

---

[1]The Royal MENCAP Society is a UK based charity for individuals with learning disabilities.

participants listed in table 1 and 17 athletes from different sports disciplines who had been invited to MENCAP sports events. Criterion validity was examined using the scores of the SON-R and the Wechsler intelligence test (Wechsler Intelligence Scale for Children or Wechsler Adult Intelligence Scale depending on the age of the participant) of 8 table tennis players (5 males, 3 females; age range 18 - 45).

Athletes and their parents or carers had been given information about the purpose of the study and gave informed consent prior to the assessment (see appendix A). For all studies, ethical approval had been obtained from the Loughborough Ethics Committee (see appendix B). The studies included participants with different causes for intellectual disability as specific causes of the disability can only be established in a minority of cases (Kaski, 2009).

2.1.2. Instruments pilot project and studies 1 to 5

SON-R 5½-17 intelligence test

The degree of intellectual functioning was measured using the shortened version of the non-verbal intelligence test SON-R (Snijders, Tellegen & Laros, 1989) as described in more detail in section 1.2.3. It consists of the following four subtests: 'Categories', 'Analogies', 'Situations' and 'Mosaics'. The subtests 'Categories' and 'Analogies' assessed abstract reasoning abilities, 'Situations' tested concrete reasoning and 'Mosaics' assessed visuo-spatial abilities.

For the subtest 'Categories' the participant was shown three drawings of objects that were related to each other. The participant was then asked to choose from a set of five more drawings those two that belonged with the first three. The subtest 'Analogies' required the participant to discover the principle behind the transformation of a geometrical figure and to apply that change to another geometrical figure. For the subtest 'Situations', the participant was

shown a drawing of which one or more parts were missing. The participant had to choose the correct part from a number of alternatives in order to complete the picture in a coherent way. For the subtest 'Mosaics', the participant was asked to copy various mosaic patterns from a picture into a two-dimensional frame using red, white and patterned squares.

All tests were administered according to the instructions of the SON-R 5 ½ - 17 Manual and Research report (Snijders, Tellegen & Laros, 1989). The only deviation from the instructions concerned the feedback towards the participants following the completion of each item: instead of giving a negative feedback when the participant gave an incorrect answer, the examiner just said 'okay'. The feedback after a correctly scored item was 'well done' instead of 'that's correct'. The terms 'correct' and 'incorrect' were avoided in order to prevent the participants from getting discouraged when answering items incorrectly. IQ scores were calculated using SON-R software provided by Hogrefe publisher.

ABC physical aptitude test

Physical performance was evaluated using the ABC physical aptitude test battery which measured a combination of Agility, Balance and Co-ordination skills. ABC is the basis of the FUNdamentals programme of physical activity development devised by Istvan Balyi of the Pacific Sport Vancouver and National Coaching Institute in Victoria, British Columbia, Canada. The ABC principles are embedded in many training methods of professional sports teams and those of National Governing Bodies of Sport/Canada. Therefore, the equipment was familiar to some as the training methods used by their sporting heroes, which helped to encourage and motivate the participants.

The test battery consisted of six different tests using the Davies Agility Drills equipment. Prior to testing, participants were familiarized with the equipment and test procedures were explained. Then, the test assistant demonstrated the tests, and, subsequently, each participant had a trial run. Participants were asked to complete the tasks as accurately as possible and to focus on their agility, balance and co-ordination, rather than speed. After the trial runs, the participant would have a short break and then start again. This time the results would be recorded.

For the first test (bunny jumps), participants had to jump with both feet over eight hurdles. For the second test (double foot run-barriers), participants had to run over eight hurdles placing both feet alternately between each hurdle. For the third test (double foot run-ladder), participants had to run through the ladder placing the right followed by the left foot in each square. For the fourth test (mixed drill), participants had to start running through the first two squares placing both feet alternately between the squares and to change after a right-angled turn into a lateral run still facing the same way and placing both feet alternately between the squares. For tests one to four, the number of hurdles or rungs of ladder which the participant crossed in 30 seconds were recorded. For the fifth test (colour compass), participants had to move out-and-back between a central marker disc and four differently coloured discs arranged in a square, equidistant from the centre (5 metres). On hearing each colour, the subject moved as quickly as possible to the nominated disc and then returned to the centre disc. On arriving there, the tester called out another colour, and the subject travelled to and from this target disc, and so on until the time elapsed (30 seconds). Performance was assessed by recording the distance (in meters) the participant had run within the time limit. The final test (chicane) was timed and combined elements of the previous five tests. Participants had to negotiate a short course that included stepping in a ladder, with a lateral change of direction, double foot jumps over low barriers, and a 'slalom' course that involved changes of

63

direction. The participants started with a walking pace trial which was followed by a timed run.

In addition to the total distance, the number of errors was also recorded for each test. Errors were noted when the participant was over-balancing in any direction to a degree that the progress was affected; lost co-ordination; touched the hurdles or ladder rungs; performed single steps instead of double steps; became confused or made any other obvious faults.

A preliminary study using two test assistants supported test-retest reliability of the ABC mean score for athletes with intellectual disabilities (r_s= .70, p< .01, n=16) and inter-rater reliability of the ABC error score (r= .66, p< .01, n=27) for athletes with intellectual disabilities. Inter-rater reliability for ABC mean was not established as performance was based on recording the distance and time. However, results did not support test-retest reliability of the ABC error score for athletes with ID (r_s = .15; p = .61) which was considered in the interpretation of the findings.

2.1.3. Test environment and procedures

All tests were administered by trained test assistants, and participants also had ample time to familiarize themselves with the test environment. The assessment with the SON-R took place in quiet, large rooms with sufficient space between testing stations to ensure that participants could be tested simultaneously without disturbing each other's performance.

Pilot Project: Recreational football players

All testing took place on the County Football Association sports grounds in Leicester on two different days. Eight participants were tested on each day. The same examiners administered the same tests on both occasions. The

64

SON-R 5 was split into two parts. One examiner (author of this thesis) administered the subtests 'Categories' and 'Mosaics', the other examiner (psychology researcher) the subtests 'Situations' and 'Analogies'. Subtest order was fully order-balanced to eliminate possible order effects. The participants had a break between each test component. Each examiner tested four participants in the morning and four were tested in the afternoon, with a lunch break in between. The ABC physical aptitude test was conducted in an enclosed and quiet sports field outside the building. No non-test related physical exercise was undertaken during the testing.

Study 1: Track and field athletes

The assessment was conducted on two different occasions at the English Institute of Sport in Sheffield during national training days. On day one, 6 participants were tested and the SON-R was split into two parts: One examiner (author of this thesis) administered the subtests 'Categories' and 'Mosaics', the other examiner (trained final year psychology student) the subtests 'Situations' and 'Analogies'. On day two, 11 participants were tested and the intelligence test was split into four parts. Each subtest was administered by a different examiner (author of this thesis, one psychology research student and two final year psychology students). The ABC physical aptitude test was administered in the back of the sports hall.

Due to the training schedule, participants had physical exercise before the testing. However, all participants had ample rest before testing started and did not exercise between the administrations of the different subtests.

Study 2: Table tennis players

Participants were tested on three different occasions during national training camps at Meres Leisure Centre in Grantham. Depending on the number of

participants being tested on the day, the SON-R was split into two or four parts with examiners administering either one or two subtests. Table 2 shows numbers of participants and examiners on each testing day.

Table 2 *Number of participants and examiners in study 2*

	Number participants	Number examiners**	Number of subtests per examiner
Testing 1	8	4	1
Testing 2*	8	4	2
Testing 3*	4	2	1
Testing 4*	4	4	1

*Testing 2, 3 and 4 also included participants for studies 3, 4 and 5. These participants are not included in table 2, but described in the following sections.

**One examiner in each testing session was the author of this thesis, the others were trained final year and/or research psychology students.

On testing days 1 and 2 the ABC physical aptitude test was administered in a separate sports hall, and on testing days 3 and 4, a separated part of the table tennis hall was used for testing. Panels around the testing area ensured that distractions were kept to a minimum for the participants. Due to the training schedule of the athletes, participants had physical exercises before the testing. However, all participants had ample rest before testing commenced and they did not participate in training between the administrations of the different subtest sessions.

Study 3: Swimmers

Participants for study 3 were tested on two different occasions. 14 participants were tested during a national swimming competition at Forge's

Pond swimming venue in Sheffield and 4 participants were tested at Meres Leisure Centre in Grantham. On both occasions the intelligence test was split into four parts. Each subtest was administered by a different examiner (the author of this thesis and trained final year psychology students). The ABC physical aptitude test was administered in a separate sports hall. The participants tested at Meres Leisure Centre in Grantham were assessed on testing day 3 in study 2 (see: test environment and procedures study 2).

Study 4: Reliability of the SON-R for individuals with intellectual disabilities

Study 4 included 14 participants who had been assessed for study 2 during table tennis training camps at Meres Leisure Centre in Grantham. For the test-retest study 7 participants were retested with the SON-R by two test assistants. Five of them were retested four months later by the same test assistants. Two were retested six months later by the same test assistants. Table 3 shows numbers of participants, examiners and subtests per examiner for initial assessment and retesting.

Table 3 Number of participants and examiners in retest -study

	Number participants	Number examiners*	Number of subtests per examiner
First assessment	7	2	2
Retesting after 4 months	5	2	2
Retesting after 6 months	2	2	2

*One examiner in each testing session was the author of this thesis, the other was a trained final year student.

For the inter-rater reliability study 7 participants were retested with the SON-R by four test assistants six weeks after the first assessment. The intelligence test was divided into its four subtests. Each subtest was administered by a different test assistant. For the retesting session, the tests were interchanged between the test assistants. Table 4 shows numbers of participants, examiners and subtests per examiner for the initial assessment and retesting.

Table 4 Number of participants and examiners in inter-rater study

	Number participants	Number examiners*	Number of subtests per examiner
First assessment	7	4	1
Retesting after 6 weeks	7	4	1

*One examiner in each testing session was the author of this thesis, the other was a trained final year student.

For both studies, test-order was order balanced to eliminate possible effects of test sequence. Due to the training schedule of the athletes, participants had physical exercises before the testing. However, all participants had ample rest before testing commenced, and they did not participate in physical exercise between the administrations of the different subtests of the SON-R.

Study 5: Validity of the SON-R for individuals with intellectual disabilities

Data for study 5 was obtained in the pilot project, in studies 1 to 3 (for procedures of data collection see chapter 2.1.3.) and in three additional testing sessions for MENCAP sport events at a college and Lee Valley

Athletics Centre. For the testing session at the college, all subtests were administered by the same test assistant (author of this thesis). For the testing session at Lee Valley Athletics Centre, the intelligence test was divided into two parts. One test assistant (author of this thesis) administered the subtests 'Situations' and 'Analogies' and another test assistant (trained psychology research student) administered the subtests 'Mosaics' and 'Categories'. On these testing occasions, the participants did not engage in physical exercise before or during the test administration.

2.1.4. Statistical Analysis

Pilot project and studies 1 to 3: Recreational football, track and field athletics, table tennis and swimming

The degree of intellectual disability was measured by using the four scores of the subtests 'Categories', 'Analogies', 'Situations' and "Mosaic". The analysis was computed using the mean intelligence score (IQ score) as well as the raw scores of the four subtests separately. The IQ score was derived from the scores of the subtests using the software provided by Hogrefe publishers.

For the ABC physical performance test each participant received two scores:

1) The total physical errors score, which was the sum of all errors made during the physical aptitude test.
2) The mean physical performance score, which was calculated as the mean of all subtest physical performance scores, using the time of the timed subtest as a negative value (i.e. mean = (test1 + test2 + ... + test5 – time(test6))/6).

The ABC scores allowed an overall analysis of physical performance of all athletes. For the analysis of the relationship between intellectual functioning and sports performance in individual sports disciplines, sport performance scores were used for table tennis and swimmers. Track and field athletes did not have sport performance scores as this sample included track as well as field athletes. These disciplines could not be combined in a linear mixed effects model and due to the small number of participants separate models could not be calculated.

- Performance scores for table tennis players were calculated using the results of two national table tennis competitions in the category 'intellectual disability' (MENCAP Sport National Championship 2006 and MENCAP Grand Prix 2007). Scores of all sets played by a participant in these two competitions were added up and the total was divided by the number of sets played by each player.

- Performance scores for swimmers were computed using final times of a national-level swimming championship to construct a linear mixed effects model accounting for the speed of each swimmer, whilst taking into account swimming style, distance, distance^2 (to model non-linear effects of distance on speed), age and gender. The model provided a performance score for each swimmer based on all final times, which was used as the swimming performance outcome variable.

A hierarchical linear regression analysis was computed to establish the association between cognitive abilities and physical performance for all participants as well as separately for participants with an IQ of 75 and below, which would be the target group of the IPC for the Paralympics.

Subsequently, the data was stratified in order to explore the association between cognitive abilities as measured with the SON-R intelligence test, and physical performance, for different sports disciplines separately. Hierarchical (stepwise) linear regression analyses were conducted with physical performance scores (error score and mean score) as dependent variables, using overall IQ scores, as well as the scores on the

SON-R subtests separately as independent variables, in order to find which tests predicted physical performance most accurately, while controlling for age and sex.

For all studies, descriptive statistics were calculated and assumptions were tested. For regression analyses, the assumption of multicollinearity was assessed using the values for tolerance and variance inflation factor (VIF). Assumptions were met when values for tolerance were higher than .10 and for VIF less than 10 (Pallant, 2005). The presence of outliers was determined by an examination of standardized residuals, with values above 3.3 and below -3.3 being identified as outliers (Tabachnick & Fidell, 2007). In an examination of the distribution of standardized residuals, normality, linearity and homoscedasticity was explored. A rectangular distribution shape of the residuals with the majority of residuals concentrated along zero would confirm normality, linearity and homocedasticity (Tabachnick & Fidell, 2007).

Power was calculated retrospectively for the regression analyses as no prior research had established observed effect sizes or variances that could be used to calculate power a priori. For the evaluation of results, a power level of 0.80 was regarded as sufficient (Pallant, 2005; Field, 2005). A level of significance of 0.05 was used (two-sided). All analyses were conducted in SPSS 14.0.

Study 4 and 5: Evaluation of reliability and validity of the SON-R for individuals with intellectual disabilities

Test-retest and inter-rater reliability of the SON-R for individuals with ID were estimated using scores on the four subtests separately, as well the overall IQ score, which was calculated with the SON-R software provided by Hogrefe publisher. Test-retest and inter-rater reliability were estimated based on the scores of the subtests obtained in the first and second testing session using Spearman's rank correlations.

71

Construct validity of the SON-R for the assessment of individuals with ID was examined using a exploratory principal component analysis on the scores of the four subtests 'Categories', 'Mosaics', 'Situations' and 'Analogies'.

In order to establish criterion validity for the SON-R for the assessment of individuals with ID, Spearman's rank correlations were calculated between IQ scores on the SON-R and registration IQ scores. The participants had obtained these scores as part of the registration procedure for the English table tennis team in the category 'Intellectual Disability'.

Prior to all analyses, descriptive statistics were computed for all analysis and assumptions were tested. For an exploratory principal component analysis assumptions include sample size, linearity of relations, factorability of relations and absence of outliers (Pallant, 2005). Required sample size depends on number of variables included in the analysis and should exceed five cases per variable but include at least 150 participants (Pallant, 2005). However, if correlations between variable are high (more than 0.6), a sample size of 100 is adequate (Field, 2005). In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was assessed in order to assess if the sample size was sufficient. The KMO should exceed 0.6 (Kaiser, 1974). Linearity of relations between the subtests was assessed based on inspection of the scatterplots (Pallant, 2005). Factorability of relations was examined by an inspection of the correlation matrix of the variables and using Bartlett's Test of Sphericity. Correlation coefficients of .3 and above are considered the minimum strength for inter-correlations (Tabachnick & Fidell, 2007). A statistically significant result of Bartlett's Test of Sphericity indicates factorablility of relations (Field, 2005).

The level of significance was set at 0.05 (two-sided). All data were analysed using SPSS version 14.0.

2.2. Results

2.2.1. Pilot project and studies 1 to 3: Recreational football, track and field athletics, table tennis and swimming

Descriptive statistics

The association between cognitive abilities and physical performance was first investigated for all participants included in the pilot project and studies 1 to 3 who completed the SON-R and the ABC physical aptitude test. Initially, the data set included 70 participants (51 male, 19 female) between 8 and 50 years of age, with a mean age of 22.81 and a standard deviation (SD) of 9.38. IQ scores ranged from 48 to 110, with a mean score of 64.09 and a SD of 15.92. All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

The ABC mean scores ranged from 1.03 to 30.88, with a mean score of 18.66 and a SD of 6.64. All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

The ABC error scores ranged from 2 to 82, with a mean score of 18.61 and a SD of 12.72. The ABC error score outlier had a value of 82, which was more than 3.29 SD outside the mean and more than 2 SD higher than the next highest score. An examination of the data collection log book revealed that this participant had severe physical coordination problems. These might be due to his cognitive impairment or could stem from an unrelated physical condition. Therefore, it was decided to compute the analysis with and without this participant.

A multiple hierarchical regression analysis was conducted to examine the association between ABC mean and IQ scores while controlling for sex and age. Results showed that IQ scores predicted 40% of the variance in ABC mean (beta= .50, $p < 0.01$, n=70) when controlling for sport discipline, sex and

age. Table 5 shows the model summary of the hierarchical regression analysis for ABC mean scores.

Table 5 Summary of Hierarchical Regression Analysis for ABC mean (N = 70) – overall sample

Variable	B	SE B	β	p
Step 1				
Age	- .28	.08	- .39	.00**
Sex	- 3.13	1.63	- .21	.06
Sport discipline	.63	.69	.10	.37
Step 2				
Age	- .13	.08	- .18	.09
Sex	- 2.49	1.44	- .17	.09
Sport discipline	-.24	.63	-.04	.71
IQ score	.21	.05	.50	.00**

Note. $R^2 = .21$ for Step 1; $\Delta R^2 = .19$ for Step 2 (ps<.01).
** p< .01

As described in section 2.1.4., the assumptions for hierarchical regression analysis were assessed: tolerance and variance inflation factor (VIF) were checked and indicated that assumptions for multicollinearity were met. An examination of the distribution of standardized residuals confirmed normality, linearity, homoscedasticity and the absence of multivariate outliers (Tabachnick & Fidell, 2007). A post-hoc power analysis for the regression analysis predicting the ABC mean scores revealed an observed power of 0.99, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

A hierarchical regression analysis revealed that a model including IQ scores and controlling for sex, age and sports discipline also showed a significant association with ABC error scores (R^2 = 17%, beta= -.22, p <.05, n=70). However, IQ score was not a significant predictor in that model. Table 6 shows the summary of the hierarchical regression analysis for ABC error scores when including the outlier.

Table 6 Summary of Hierarchical Regression Analysis for ABC error including outlier (N = 70) – overall sample

Variable	B	SE B	β	p
Step 1				
Age	.42	.16	.31	.01
Sex	- 4.51	3.25	- .16	.17
Sport discipline	-1.71	1.38		
Step 2				
Age	.30	.17	.22	.08
Sex	- 5.04	3.23	-.18	.12
Sport discipline	-1.00	1.42	-.08	.48
IQ score	-.17	.10	-.22	.10

Note. R^2 = .14 for Step 1; ΔR^2 = .04 for Step 2 (ps>.05).

Assumptions for hierarchical regression analysis were examined as described in chapter 2.1.4.: Tolerance and variance inflation factor (VIF) were both checked and indicated that assumptions for multicollinearity were met. An examination of the distribution of standardized residuals confirmed normality, linearity, homoscedasticity and also identified the outlier (Tabachnick & Fidell, 2007). For the regression analysis predicting the ABC error score, the observed power was 0.85 when including the outlier, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

75

When computing the hierarchical regression analysis without the outlier, the association between ABC error scores and IQ scores was trend significant (R^2 = 14%, beta = -.272, p = .05, n = 69) when controlling for sport discipline, sex and age with the IQ score being the only significant predictor in the model. Table 7 shows the summary of the hierarchical regression analysis for ABC error scores when excluding the outlier.

Table 7 Summary of Hierarchical Regression Analysis for ABC error excluding outlier (N = 69) – overall sample

Variable	B	SE B	β	p
Step 1				
Age	.15	.14	.13	.27
Sex	- 3.13	2.71	-.14	.25
Sport discipline	-1.99	1.14	-.21	.09
Step 2				
Age	.03	.15	.03	.83
Sex	- 3.66	2.66	-.16	.17
Sport discipline	-1.28	1.17	-.13	.28
IQ score	-.17	.09	-.27	.05*

Note. R^2 = .08 for Step 1; ΔR^2 = .06 for Step 2 (ps<.05). * p<.05

When the analysis was repeated only including participants with an IQ score of 75 or below, the results confirmed the association for that population. Results showed a significant association between IQ scores and ABC mean scores (R^2= 37%, beta = .47, p< .01, n= 56[1]), as well as between IQ scores and ABC error scores (R^2= 28%, beta = -.32, p< .01, n= 56), when controlling for sex, age and sports discipline. Again, without the outlier the strength of the

---

[1] 56 of the 75 participants had an IQ score of 75 or below

association between IQ scores and ABC error scores (R^2= 21%, beta = -.34, p< .05, n= 55) dropped slightly.

These findings supported the alternative hypothesis that there is a positive association between IQ scores and physical performance for elite athletes with ID.

In order to investigate the association between the different cognitive abilities, as measured in the different subtests, and physical performance for the different sports disciplines separately, descriptive statistics have been computed per sports discipline. Descriptive statistics included the variables age, IQ score, scores on the subtests 'Categories', 'Mosaics', 'Situations' and 'Analogies', ABC error scores, ABC mean scores and competition performance scores for table tennis players and swimmers. The floor effect in the IQ score was not necessarily caused by poor discrimination of all subtests: therefore, distributions of standardized residuals were checked for each study individually.

Pilot project: recreational football players

The pilot project included recreational football players (11 male, 4 female). Table 8 shows minimum and maximum scores, mean and SD for the participants of the pilot project.

Table 8 Descriptive statistics pilot project (n=15) – football players

	Minimum	Maximum	Mean	SD
age	17	23	20.88	1.61
IQ scores	48	83	59.87	11.04
'Categories' scores	0	17	6.53	4.93
'Mosaics' scores	0	14	6.27	4.18
'Situations' scores	0	23	10.47	7.30
'Analogies' scores	1	21	9.13	6.32
ABC error	2	45	20.53	10.47
ABC mean	1.59	22.79	16.51	6.02

All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

The results of a hierarchical regression analysis revealed that overall IQ scores showed only a trend significant association with ABC *error* scores when controlling for sex and age (R^2 adj. = .30, beta = -.68, p= .08) with IQ scores being the only significant contributor to the model. However, when entering the subtests separately, a stepwise hierarchical regression analysis showed that a model using scores on the subtest 'Analogies' significantly predicted ABC error scores (R^2 adj.= .53, beta = -.73, p<.05) when controlling for sex and age. Table 9 shows the model summary of the hierarchical regression analysis for ABC error scores when none of the other subtests was entered in the analysis.

Table 9 Summary of Hierarchical Regression Analysis for ABC error scores (N = 15) – football players

Variable	B	SE B	β	p
Step 1				
Age	- 2.55	1.83	-.39	.19
Sex	8.04	6.45	.35	.24
Step 2				
Age	- 1.23	1.33	-.19	.37
Sex	1.57	4.84	.07	.75
subtest 'Analogies'	-1.21	.33	-.73	.00**

Note. R^2 = .18 for Step 1; ΔR^2 = .45 for Step 2 (ps<.01).

** p< .01

As discussed in section 2.1.4. assumptions for hierarchical regression analysis were assessed: tolerance and VIF were checked and indicated that assumptions for multicollinearity were met. An inspection of the standardized residual scatter plot confirmed normality, linearity, homoscedasticity and the absence of multivariate outliers (Tabachnick & Fidell, 2007). A post-hoc power analysis for the regression analysis predicting the ABC error scores revealed an observed power of 0.83, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

.

When using stepwise hierarchical regression analysis with the subtests separately, and controlling for sex and age, ABC mean scores were not shown to be significantly associated with any of the subtests or the IQ score.

The results of a Spearman's rank correlation confirmed the relationship between physical performance and cognitive abilities for this group of recreational football players: ABC error scores showed significant correlations with all SON-R subtests, while ABC mean was only significantly associated with the subtest 'Categories' (table 10).

Table 10 Spearman's rank correlation between ABC error scores and SON-R subtests (N = 15) – football players

	IQ scores	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'	Subtest 'Analogies'
ABC error	-.76**	-.67**	-.55*	-.61 *	-.77**
ABC mean	.47	.55*	.48	.34	.27

* p< .05; ** p<.01

Study 1: Track and field athletics

Study 1 included track and field athletes (14 male, 3 female). Table 11 shows minimum and maximum scores, mean and SD.

Table 11 Track and field athletics (N=17)

	Minimum	Maximum	Mean	SD
age	14	40	19.82	7.18
IQ scores	48	95	63.41	11.41
'Categories' scores	3	21	8.88	3.97
'Mosaics' scores	0	13	6.41	3.73
'Situations' scores	3	27	12.71	5.79
'Analogies' scores	0	24	11.12	6.17
ABC error scores	2	33	18.94	9.23
ABC mean scores	13.87	30.88	22.46	4.93

All values were within 3.29 SD of the mean and therefore no outliers were identified (Tabachnick & Fidell, 2007).

The results of a hierarchical regression revealed that there was no significant association between the overall IQ scores or the SON-R subtest scores and score on the ABC physical aptitude test when controlling for sex and age. This was confirmed for ABC error scores with the results of a Spearman's rank correlation, which showed no correlations between subtests and ABC error scores. However, the results of the Spearman's rank correlation did indicate a relationship between ABC mean and overall IQ scores (r_s= .50, p<.05, n=17) as well as scores on the subtest 'Situations' (r_s= .50, p<.05, n=17).

Study 2: Table tennis

Participants for study 2 were table tennis players (16 male, 8 female). Table 12 shows minimum and maximum scores, mean and SD.

Table 12 Table tennis (N=24)

	Minimum	Maximum	Mean	SD
Age	15	50	29.54	10.60
IQ scores*	48	106	60.65	15.72
'Categories' scores	1	22	6.88	5.10
'Mosaics' scores	0	19	6.21	5.11
'Situations' scores*	3	26	12.35	6.79
'Analogies' scores	0	28	7.42	8.55
ABC error scores	2	82	19.58	16.88
ABC mean scores	1.03	29.89	15.71	6.79
Competition score**	1.93	10.76	7.86	2.41

Note. * One participant did not have a score on the subtest 'Situations' and therefore no overall IQ score could be calculated for this participant; ** n=18

Except for ABC error scores, all values were within 3.29 SD of the mean and, therefore no outliers were identified (Tabachnick & Fidell, 2007). The ABC error score outlier had a value of 82, which was more than 3.29 SD outside the mean and more than 2 SD higher than the next highest score. The sample for table tennis included the participant with severe physical coordination problems. Therefore, it was decided to compute the analysis with and without this participant.

A hierarchical regression analysis revealed that IQ scores was significantly associated with ABC mean scores when controlling for sex and age scores (R^2 adj. = .39, beta = .63, p<.01). Table 13 shows the model summary of the hierarchical regression analysis for ABC mean scores.

Table 13 Summary of Hierarchical Regression Analysis for ABC mean scores using IQ scores – table tennis (N = 23)

Variable	B	SE B	β	p
Step 1				
Age	- .29	.13	-.45	.04*
Sex	-1.46	2.80	-.10	.61
Step 2				
Age	- .07	.13	-.10	.91
Sex	- .29	2.41	.02	.91
IQ scores	.27	.09	.63	.01**

Note. R^2 = .21 for Step 1; ΔR^2 = .27 for Step 2 (ps<.01).

Assumptions for hierarchical regression analysis were examined as described in section 2.1.4.: tolerance and VIF were checked and supported assumptions for multicollinearity. An inspection of the standardized residuals confirmed normality, linearity, homoscedasticity and the absence of multivariate outliers

(Tabachnick & Fidell, 2007). A post-hoc power analysis for the regression analysis predicting the ABC mean scores revealed an observed power of 0.84, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

When entering the subtests separately, and controlling for sex and age, a stepwise hierarchical regression analysis showed that a model using scores on the subtest 'Categories' significantly predicted ABC mean scores (R^2 adj.= .44, beta = .66, p<.01). Table 14 shows the model summary of the hierarchical regression analysis for ABC mean scores.

Table 14 Summary of Hierarchical Regression Analysis for ABC mean scores using scores on the subtest 'Categories' – table tennis (N = 24)

Variable	B	SE B	β	p
Step 1				
Age	-.29	.13	-.45	.04*
Sex	-1.46	2.80	-.10	.61
Step 2				
Age	-.06	.12	-.10	.60
Sex	-.87	2.25	-.06	.70
Subtest 'Categories'	.87	.25	.66	.00**

Note. R^2 = .21 for Step 1; ΔR^2 = .31 for Step 2 (ps<.01). *p<.05, **p< .01

Assumptions for hierarchical regression analysis were investigated as described in section 2.1.4: tolerance and VIF were checked and supported assumptions for multicollinearity. An inspection of the standardized residuals confirmed normality, linearity, homoscedasticity and the absence of multivariate outliers (Tabachnick & Fidell, 2007). A post-hoc power analysis for the regression analysis predicting the ABC mean scores revealed an

observed power of 0.93, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

.

A hierarchical regression analysis showed that overall IQ scores were not significantly associated with table tennis competition scores (R^2 adj.= .17, beta = .55, p = .12, n= 19). However, using a hierarchical linear regression analysis and entering all subtests separately, a model using the subtest 'Categories' predicted table tennis competition scores (R^2 adj.= .30, beta = .66, p<.05, n= 19) when controlling for age and sex. Table 15 shows the model summary of the hierarchical regression analysis for table tennis competition scores.

Table 15 Summary of Hierarchical Regression Analysis for table tennis competition scores using scores on the subtest 'Categories' – table tennis (N = 19)

Variable	B	SE B	β	p
Step 1				
Age	- .02	.05	-.09	.70
Sex	-1.61	1.18	-.32	.19
Step 2				
Age	.06	.05	.26	.29
Sex	- 1.40	.99	-.28	.18
Subtest 'Categories'	.31	.11	.66	.01**

Note. R^2 = .11 for Step 1; ΔR^2 = .30 for Step 2 (ps<.05). *p< .05

As described in section 2.1.4., assumptions for hierarchical regression analysis were assessed: tolerance and VIF were checked and supported assumptions for multicollinearity. An inspection of the standardized residuals confirmed normality, linearity, homoscedasticity and the absence of

multivariate outliers (Tabachnick & Fidell, 2007). A post-hoc power analysis for the regression analysis predicting the table tennis competition scores revealed an observed power of 0.54, which is below the desired level of power (Pallant, 2005; Field, 2005). The consequences of this low level of power will be considered in the discussion.

None of the SON-R subtests predicted ABC error scores while controlling for age and sex. The results were independent of in- or exclusion of the outlier.

When using Spearman's rank correlation, results confirmed the association between physical performance and cognitive abilities for ABC mean scores as well as for table tennis competition scores: IQ scores were significantly associated with ABC mean scores ($r = .57$, $p < .01$, $n = 23$) and with table tennis competition scores ($r = .48$, $p < .05$, $n = 19$), but not with ABC error scores (independent of the in- or exclusion of the outlier). ABC mean scores were significantly associated with all subtests. Table tennis competition scores were significantly associated with the subtests 'Categories' and 'Mosaics' (table 16).

Table 16 Spearman's rank correlations between ABC mean/ table tennis competition scores and SON-R subtests

	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'	Subtest 'Analogies'
ABC mean (N= 24)	.64**	.55**	.43*	.49**
Competition scores (N=19)	.59**	.55*	.23	.37

Note. Only 23 participants had scores on the subtest 'Situations'; * p< .05; ** p<.01

Study 3: Swimming

Participants for study 3 were elite swimmers (13 male, 5 female). Table 17 shows minimum and maximum scores, mean and SD.

Table 17 Swimming (N=18)

	Minimum	Maximum	Mean	SD
Age	8	40	18.67	8.21
IQ scores	48	110	74.94	21.84
'Categories' scores	0	22	10.72	8.00
'Mosaics' scores	0	20	10.06	5.90
'Situations' scores	2	27	14.44	8.01
'Analogies' scores	0	28	7.42	8.55
ABC error scores*	2	38	14.33	10.03
ABC mean scores*	9.03	29.29	21.02	5.92
Competition score**	-7.64	6.33	0.00	3.66

Note: * n=15; ** n=12

All values were within 3.29 SD of the mean and therefore no outliers were identified (Tabachnick & Fidell, 2007).

The results of a hierarchical regression analysis revealed a trend significant association between overall IQ scores and ABC mean scores (R^2 adj.= .30, beta = .66, p = .08). However, when entering the subtests separately into the analysis, and controlling for sex and age, a stepwise hierarchical regression analysis showed that the subtest 'Mosaics' significantly predicted ABC mean scores (R^2 adj.= .56, beta = .79, p<.01). Table 18 shows the model summary of the hierarchical regression analysis for ABC mean scores.

86

Table 18 Summary of Hierarchical Regression Analysis for ABC mean scores using scores on the subtest 'Mosaics' – swimming (N = 15)

Variable	B	SE B	β	p
Step 1				
Age	- .03	.20	-.04	.89
Sex	-3.18	3.59	-.25	.39
Step 2				
Age	- .03	.13	-.04	.81
Sex	-.70	2.35	-.05	.77
Subtest 'Mosaics'	.80	.18	.79	.00**

Note. R^2 = .06 for Step 1; ΔR^2 = .59 for Step 2 (ps<.01). **p< .01

Assumptions for hierarchical regression analysis were examined as described in section 2.1.4.: tolerance and VIF were checked and supported assumptions for multicollinearity. The inspection of the standardized residuals supported the assumptions of normality, linearity and homoscedasticity for ABC mean scores.

A post-hoc power analysis for the regression analysis predicting the ABC mean scores revealed an observed power of 0.88, which is within the acceptable parameter (Pallant, 2005; Field, 2005).

The results revealed no significant association between scores on the SON-R subtests and ABC error scores or swimming competition scores when using a hierarchical linear regression analysis.

A Spearman's rank correlation supported these results: Overall IQ scores were significantly associated with ABC mean scores (r_s= .52, p<.05, n=15) while ABC error scores and swimming competition scores were not associated with IQ scores.

When looking at the SON-R subtests separately using a Spearman's rank correlation, all subtests (except the subtest 'Categories') were significantly associated with ABC mean scores (table 19). The subtest 'Categories' was only trend significant (r_s=.50, p= .06, n=15). For ABC error scores, only the subtest 'Mosaics' showed a significant correlation (r_s= -.54, p< .05, n=15) when using a Spearman's rank correlation.

Table 19 Correlations using Spearman's rho for SON-R subtests separately – swimming (N=15, two-tailed)

	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'	Subtest 'Analogies'
ABC mean	.50	.83**	.54*	.62*

* p< .05; ** p<.01

2.2.2. Studies 4 and 5: Reliability and validity of the SON-R 5½-17 intelligence test for individuals with intellectual disabilities

Study 4: Reliability of the SON-R 5½-17 intelligence test for individuals with intellectual disabilities

Test-retest reliability

The assessment of test-retest reliability of the SON-R 5½-17 intelligence test for individuals with ID included 7 participants (5 male, 2 female) with an age range between 31 and 50 years (mean = 38.14, SD = 6.77). Table 20 shows minimum and maximum scores, mean and SD.

Table 20 Test-retest reliability (n=7)

	Minimum		Maximum		Mean		SD	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest
IQ scores	48	48	68	79	53.43	56.43	7.70	11.79
'Categories'	2	0	10	13	5.00	5.71	3.22	5.41
'Mosaics'	0	0	7	10	4.14	4.00	2.34	3.27
'Situations'	3	2	22	24	8.57	10.00	6.97	7.66
'Analogies'	0	0	11	19	3.71	6.29	4.46	7.20

Due to the small number of participants in this study, a Spearman's rank correlation was computed to establish test-retest reliability. Therefore, it was not necessary to assess assumptions of normality.

A Spearman's rank correlation confirmed test-retest reliability of the SON-R 5½-17 intelligence test for individuals with ID for the overall IQ score (r_s=.88, p<.01, n=7) as well as for the subtests separately (see table 21).

Table 21 Test-retest reliability using Spearman's rho for SON-R subtests separately (N= 7, two-tailed)

	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'	Subtest 'Analogies'
Spearman's rho	.76*	.96**	.96**	.90**

* p< .05 ** p<.01*

Inter-rater reliability

The assessment of inter-rater reliability of the SON-R 5½-17 intelligence test for individuals with ID also included 7 participants (5 male, 2 female) with an age range between 17 and 42 years (mean age = 27.29, SD = 9.96) . Table 22 shows minimum and maximum scores, mean and SD.

Table 22 Inter-rater reliability (N = 7)

	Minimum		Maximum		Mean		SD	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest
IQ scores	49	49	72	71	57.14	61.86	7.97	9.82
'Categories'	1	4	11	12	6.57	7.43	3.87	2.94
'Mosaics'	3	5	11	17	6.43	8.57	2.88	4.16
'Situations'	6	1	18	25	11.57	13	3.82	7.74
'Analogies'	0	0	22	21	5.43	7.71	7.57	7.74

Again, due to the small number of participants in this study, a Spearman's correlation was computed to establish inter-rater reliability. Therefore, it was not necessary to assess assumptions of normality.

A Spearman's rank correlation did not support inter-rater reliability of the SON-R for individuals with ID for the subtests 'Situations' and 'Analogies' (see table 23) or the overall IQ score (r_s=.67, p=.11, n=7).

Table 23 Inter-rater reliability using Spearman's rho for SON-R subtests separately (N= 7, two-tailed)

	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'	Subtest 'Analogies'
Spearman's rho	.83*	.94**	.45	.39

* p< .05 ** p< .01*

Study 5: Validity of the SON-R 5½-17 intelligence test for individuals with intellectual disabilities

Study 5 investigated construct and criterion validity of the SON-R 5½-17 intelligence test for individuals with ID.

Construct validity

Construct validity was assessed using an exploratory factor analysis of the scores on the different subtests of 91 participants (63 male, 28 female). Table 24 shows minimum and maximum scores, mean and SD.

Table 24 Construct validity (N = 91)

	Minimum	Maximum	Mean	SD
Age	8	50	23.29	9.63
'Categories'	0	22	7.80	5.89
'Mosaics'	0	20	7.40	5.61
'Situations'	0	28	12.40	7.67
'Analogies'	0	29	23.29	9.63

All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007). Assumptions were tested as discussed in section 2.1.4. Linearity of relations between the subtests were checked and confirmed based on inspection of the scatterplots. The correlation matrix revealed that all coefficients were .3 and above (table 25), which is considered the minimum strength for inter-correlations (Tabachnick & Fidell, 2007).

Table 25 Correlations between SON-R subtests (N=91)

	Subtest 'Categories'	Subtest 'Mosaics'	Subtest 'Situations'
Subtest 'Mosaics'	.73**		
Subtest 'Situations'	.74**	.76**	
Subtest 'Analogies'	.79**	.76**	.78**

** p<.01

The Kaiser-Meyer-Oklin value was .86 and therefore above the recommended value of .6 (Kaiser, 1974). Bartlett's Test of Sphericity was statistically significant. These indicators suggested that the data were suitable for factor analysis (Pallant, 2005).

The principal component analysis established one component with an Eigenvalue of 3.27, explaining 82 % of the variance (see table 26).

Table 26 Results principal component analysis (N=91)

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.27	81.96	81.96	3.28	81.96	81.96
2	.28	7.02	88.98			
3	.24	6.04	95.02			
4	.20	4.97	100.00			

The results of the principal component analysis thus showed one dominant factor, which is regarded to be the underlying intelligence factor "g" (Carroll, 1993; Tellegen & Laros, 1993). These findings supported the construct

validity of the SON-R 5½ -17 intelligence test for the use for individuals with ID.

Criterion validity

To establish criterion validity, the SON-R IQ scores of 8 table tennis players (5 males, 3 females) were compared with independently assessed Wechsler intelligence test scores (Wechsler Intelligence Scale for Children or Wechsler Adult Intelligence Scale depending on the age of the participant). Table 27 shows minimum and maximum scores, mean and SD.

Table 27 Criterion validity (n=8)

	Minimum	Maximum	Mean	SD
Age	18	45	30.38	9.97
SON-R IQ scores	48	73	58.88	9.57
Wechsler IQ scores	42	69	57.50	10.38

Due to the small number of participants in this study, a Spearman's rank correlation was computed to establish criterion validity. Therefore, it was not necessary to assess assumptions of normality.

The results revealed that scores on the SON-R and scores on the Wechsler intelligence scale were significantly correlated (r_s= .82, p< .05)[1].

The results of this analysis confirmed the criterion validity of the SON-R for individuals with ID.

---

[1] The Wechsler scores of these table tennis players, however, did not correlate with physical or sport competition scores. For further discussion, see chapter 2.3.

2.3. Discussion

The results showed a significant relationship between cognitive abilities, as assessed using the SON-R intelligence test, and physical performance for athletes with ID. Depending on the sports disciplines, different cognitive abilities were associated with physical performance.

- For table tennis players, the subtest 'Categories' was a significant predictor for both table tennis competition performance scores and scores on the ABC physical aptitude test. The subtest 'Categories' belongs to the group of non-verbal 'classification' tests assessing inductive reasoning abilities, which depend on recognition of similarities and concept formation (Mayer, 1994). This type of test is included in most commonly used intelligence tests such as Wechsler Intelligence Scales for Children (Wechsler, 2003) or the Stanford-Binet Intelligence Scale-Fifth Edition (Roid, 2003a). For table tennis, these skills are thought to be essential for the processing of information on the ball's trajectory, the likely location of the bounce and the subsequent planning of the motor response (Vickers, 1996). However, the low value for observed power in the regression analysis using table tennis competition scores indicated that the analysis should be repeated with a larger sample size to confirm the results. Furthermore, the Wechsler scores were not significantly correlated with any physical or sports competition scores while the scores on the nonverbal SON-R test showed significant correlations with ABC mean, as well as table tennis competition scores. This indicates that a nonverbal intelligence test is a better instrument to investigate the association between cognitive abilities and physical or sport performance than a conventional intelligence test.
- For swimmers, the subtest 'Mosaics' was a significant predictor for scores on the ABC physical aptitude test, but not for swimming performance. 'Mosaics' is visuo-spatial ability test based on Koh's

Block Design test (Hutt, 1932; Snijders, Tellegen & Laros, 1989) and is commonly used in intelligence tests such as the Wechsler Intelligence Scales for Children (Wechsler, 2003) or the Stanford-Binet Intelligence Scale-Fifth Edition (Roid, 2003).

- For track and field athletes, no significant association between cognitive abilities and physical performance was found. Sports performance data were not available (see section 2.1.4.)

Although results for these studies were not controlled for lifestyle issues (including BMI) or motivation, all athletes regularly visited training camps and competed at national and international sport events. Therefore, the influence of these factors was considered to be limited. However, it cannot be excluded to have affected the results. Practical limitations restricted further testing (which took place during training camps or competitions) to assess lifestyle factors.

For recreational football players with ID, the subtest 'Analogies' predicted ABC error scores. The subtest 'Analogies' is a nonverbal, culture-fair test for inductive reasoning and is similar to tests included in the Cattell Culture Fair IQ test (Cattell, 1949), which was based on the Raven's progressive matrices. This type of series completion problems requires the participant to follow a series of cognitive processes and is included in most commonly used intelligence tests such as Wechsler Intelligence Scales for Children (Wechsler, 2003) or the Stanford-Binet Intelligence Scale-Fifth Edition (Roid, 2003). However, the initial studies did not support test-retest reliability of the ABC error score for individuals with ID. This indicates that this outcome parameter may have limited predictive outcome for actual physical performance assessment. Also, the participants in the pilot study did not exercise on a regular basis. Therefore, results for this group should be confirmed using sport performance outcomes of elite-level footballers with a sport-specific physical outcome parameter (e.g. ratings by a football expert).

These studies demonstrated that the relationship between cognitive functioning and physical performance varied depending on sports discipline. While results for table tennis players and swimmers indicated a significant association between cognitive abilities and physical performance, this was not the case for track and field athletes. This sport-specific relationship was confirmed when looking at competition results. Results showed a significant association between cognitive abilities and sports competition performance for table tennis, but not for swimmers. This indicates that table tennis requires more cognitive ability than swimming.

Furthermore, not all cognitive abilities contributed equally to physical performance and, therefore, IQ in itself is inapt to determine the impact of intellectual disability on physical performance. Depending on sports discipline, different cognitive abilities predicted physical performance of athletes with ID. As discussed earlier, similar differences between sports disciplines had been found for non-disabled athletes (see chapter 1). Furthermore, the results reflected the outcome of previous studies, in that individuals with higher IQ scores showed better physical and sports performance scores (Frey & al., 1999; Van de Vliet & al., 2006).

This study had several limitations. First, there was a relatively small number of participants in each sports discipline and a wide IQ range of participants. A retrospective power analysis showed sufficient power for all regression analyses except for the analyses for table tennis competition scores using scores on the subtest 'Categories'. However, all analysis had relatively low numbers of participants and power could only be determined retrospectively (see section 2.1.4.). This could limit the generalizability of the results.

Although all participants had an intellectual disability, not all of them had been formally diagnosed based on the criteria of the British Psychological Association, the DSM-IV or the WHO, and, therefore, did not necessarily meet the criteria for the IPC target population. However, when limiting the sample to participants with IQ scores of 75 and below, results confirmed the

association between cognitive abilities and physical performance for this population.

Second, environment was controlled but interruptions (participants coming or leaving) did occur. Occasionally participants had to leave for scheduled competitions in between the subtests, but they returned after a sufficient recovery period to continue the assessment. Overall, however, participants did not seem to be unduly affected by this, and it is doubtful that higher scores would have been obtained in different environments.

Third, the age range of participants was over the range for which the tests were originally intended, but none of the participants performed at ceiling. Therefore, it was assumed that the SON-R provides enough difficulty to differentiate between IQ scores of individuals with a low degree of intellectual disability. The choice for this test was deliberate in that it had sufficient easy items. This was expected to resolve potential issues regarding low resolution towards the end of the IQ spectrum (Nunally & Bernstein, 1994). However, the fact that 13 participants received a zero score on one or more of the subtests causes some concern about the potential of the test to differentiate the IQ scores of individuals with a very high degree of intellectual disability. Additional easier items, that precede the starting items of the tests, should be added for future studies.

Validity and internal consistency reliability of the SON-R was supported by the results of this study. The results also confirmed test-retest reliability but not inter-rater reliability. However due to the small numbers of participants included in the test-retest and inter-rater reliability studies, these analyses should be repeated with more participants.

Fourth, comments from participants identified several culturally biased items in the subtests 'Categories' and 'Situations', which were based on pictorial representations. Two were included in the subtest 'Categories'. One pictured

a drawing of a continental electricity socket. These look substantially different to British sockets, as they have only two holes instead of three. Several participants seemed to be unable to recognize the object and either questioned what it was, or asked for confirmation that it was a socket. Another item that participants had difficulties to recognise, was a picture of fabric bales. Although these objects look similar in Britain, they are not commonly displayed in shops or markets and, therefore, might have been an unfamiliar sight to some of the participants. The subtest 'Situations' included another item that was questioned by the participants several times, which showed a Dutch mail box. Again, those look substantially different compared to British mail boxes. As the cultural bias studies by Tellegen and Laros (2004) had only indicated one biased item in the subtest 'Categories', the results of their study seemed not applicable for a British population.

Lastly, the range of cognitive abilities measured by the SON-R was limited to visuo-spatial insight and inductive reasoning abilities. Intellectual disability, however, might limit sports performance through other cognitive impairments which were not measured with the SON-R.

In sum, the results of these studies provide clear indications that a short nonverbal IQ test can be used to predict the physical performance of athletes with ID. Firstly, this study established the impact of the degree of cognitive functioning on sports performance. Previous research (Frey & al., 1999; Van de Vliet & al., 2006) demonstrated only a significant difference in physical performance between individuals with and without intellectual disability. This study confirmed the association between the *degree* of intellectual functioning and the level of physical performance for top-level table tennis players and swimmers, but not for track and field athletes when analysed separately. These data and the previous analysis of intelligence tests indicated that there is a clear need for a novel system to assess the degree of cognitive abilities in athletes with ID. The results also confirmed that not all cognitive abilities are

equally important for optimal sport performance, and that this was different per discipline.

Based on these results, it was decided to develop a new cognitive test battery for intellectual disability sports events. Further research with such a test battery was thus necessary to confirm the association between cognitive abilities and sports performance for athletes with ID and to validate the results for different sports disciplines. This research is described in the next part of this thesis.

Chapter III: Development and assessment of a computerized cognitive test battery

3.1. Introduction

Chapter three of this thesis is dedicated to the description of the development of a nonverbal computerized cognitive test battery and to the investigation of its psychometric properties. The purpose of the development of the CCIID was to create an instrument that is suitable for the examination of the relationship between cognitive abilities and sports performance in athletes with ID. Based on the limitations of existing verbal and nonverbal intelligence tests discussed in chapter 1 (floor effects, cultural bias), it was decided to design a novel Computerized Cognitive test battery for Individuals with ID (CCIID), which would form part of a wider test battery in order to assess a broad range of cognitive abilities.

The following chapters will present the theoretical foundation of the CCIID. I will discuss several methodological aspects that were highlighted in the previous chapters and which were, subsequently, taken into account for the construction of the test battery.

The main criteria for the test construction, which will be discussed in more detail below, were derived from theoretical considerations as well as from the results of the initial studies. These criteria included:
   a) a theoretical framework (the Cattell-Horn-Carroll theory of cognitive abilities)
   b) the results of the earlier research with the SON-R (pilot project and studies 1 to 3)
   c) the needs of the specific target group, i.e. individuals with ID

d) a focus on cultural fairness of the test and cross cultural applicability

The following chapters will discuss and analyse these criteria to develop suitable subtests for the test battery.

3.1.1. Theoretical foundation

The theoretical construct of the CCIID was based on the Cattell-Horn-Carroll theory of cognitive abilities (CHC) which is a combination of the two intelligence theories of Gf-Gc theory (McGrew, Werder & Woodcock, 1991) and Carroll's Three-Stratum Theory (Carroll, 1993). These theories were further developed and merged into the CHC in the late eighties and early nineties, which was first mentioned by McGrew, Werder and Woodcock in the Woodcock-Johnson Revised Technical Manual in 2001. The choice for the CHC was based on two aspects. Firstly, the CHC is the foundation of several commonly used intelligence tests, such as the Stanford-Binet Intelligence Scales-V, the Woodcock-Johnson Tests of Cognitive abilities-III and the Kaufman Assessment Battery for Children (Alfonso, Flanagan & Radwan, 2005). Secondly, and more in line with our findings from studies 1-3, the CHC emphasises a hierarchical structure of 10 broad and 70 narrow cognitive abilities, whereas the significance of a general intelligence factor 'g' is debatable (Davidson & Downing, 2000; McGrew, 2005).

An overall 'IQ' score was also shown in our studies not to correspond as well with physical performance, as the individual cognitive abilities, which is perhaps not surprising considering the findings of the studies presented in chapter 1, which investigated cognitive abilities of non-disabled athletes of different sports disciplines. The 10 broad abilities include: Fluid intelligence, quantitative knowledge, crystallized intelligence, reading and writing, short-term memory, visual processing, auditory processing, long-term storage and

retrieval, processing speed and decision speed/reaction time. The following sections will discuss the selection of broad abilities included in the CCIID in more detail.

3.1.2. Evaluation of pilot project and studies 1 to 3

The results of the pilot project and studies 1 to 3 showed a significant relationship between specific cognitive abilities and physical performance for elite athletes with ID. However, depending on the sports disciplines, different cognitive abilities were shown to be associated with better physical and sports performance. This, subsequently, also influenced the choice of subtests for the CCIID (see below).

For table tennis players, the SON-R subtest 'Categories' was a significant predictor for both table tennis competition performance scores as well as scores on the physical aptitude (ABC) test. 'Categories' is classified as a test for inductive reasoning abilities (Mayer, 1994), which belongs to the broad ability 'fluid reasoning'.

For swimmers, the SON-R subtest 'Mosaics' was a significant predictor for scores on the ABC physical aptitude test. 'Mosaics' is a visuo-spatial ability test, which reflects the broad ability 'visual processing' (McGrew, 2005).

For track and field athletes, none of the subtests was significantly associated with physical performance.

For recreational football players, the SON-R subtest 'Analogies' was a significant predictor for physical performance. 'Analogies' is, again, a test for inductive reasoning abilities (Mayer, 1994), which is part of the broad ability 'fluid reasoning'.

103

Consequently, the development of the CCIID thus focused on the broad abilities 'fluid reasoning' and 'visual processing', as only those abilities were shown to have a significant association with physical performance in athletes with ID. Subtests for the broad abilities short-term memory, long-term storage and retrieval, processing speed and decision speed/reaction time were included in the wider cognitive test battery, but are not part of this thesis. These abilities were considered to be out of the scope of the focus of this thesis. For the development of the CCIID test battery, verbal and general knowledge broad abilities were also excluded, as tests including those abilities are considered to be more culturally biased (Braden & Athanasiou, 2005).

The following section will discuss two of the broader abilities chosen for the CCIID in more detail.

Fluid intelligence

The broad ability 'fluid intelligence' (Gf) is defined as "the use of deliberate and controlled mental operations to solve novel, 'on-the –spot' problems. Mental operations often include drawing inferences, concept formation, classification, generating and testing hypotheses, identifying relations, comprehending implications, problem solving, extrapolating and transforming information. Inductive and deductive reasoning are generally considered the hallmark of Gf." (McGrew, 2005, p.151). In the CHC theory, Gf includes the five narrow abilities 'general sequential reasoning', 'induction', 'quantitative reasoning', 'piagetian reasoning' and 'speed of reasoning'.

Visual processing

The broad ability 'visual processing' (Gv) is defined as "the ability to generate, retain, retrieve and transform well-structured visual images" (Lohman, 1994, p.1000). Visual processing includes the 11 narrow abilities 'visualization',

'spatial relations', 'visual memory', 'closure speed', 'flexibility of closure', 'spatial scanning', 'serial perceptual integration', 'length estimation', 'perceptual illusions', 'perceptual alternations' and 'imagery'.

The broad abilities 'fluid reasoning' and 'visual processing' are measured using a variety of subtests assessing different narrow abilities. Based on the results of the pilot project and studies 1 to 3 and the theoretical background, subtests for inductive reasoning and visualization were thus chosen for the CCIID.

3.1.3. Target group

In order to gain maximum discrimination within the lower range of intelligence, it was decided that each subtest should start with the easiest possible item that represented the ability the subtest assessed and slowly increase in difficulty. Individuals with ID might perform at different ability levels in the different subtests, and, therefore, surpass the cut-off of 75 on one or more subtests, while performing below 75 on others. Consequently, it was important to include items which discriminated above an IQ of 75. In the construction of the CCIID the aim was to develop items that would discriminate from the lowest possible ability level that a particular subtest could provide, up to a level which the average population would achieve.

3.1.4. Cultural Fairness

Cultural fairness was an essential factor, given that the purpose of the test battery was to be used for elite athletes competing at international sports events. A nonverbal design was chosen to reduce cultural sensitivity, as nonverbal intelligence tests are considered less culturally biased than

intelligence tests, which also include the assessment of verbal abilities and general knowledge (Braden & Athanasiou, 2005, McCallum, 2003).

In order to further reduce cultural bias, several characteristics had to be taken into account during the development of a culturally fair test. Athanasiou (2000) constructed a framework to index test fairness for nonverbal intelligence tests which was used for the development of the CCIID. Athanasiou's criteria are grouped into three categories: i) foundation, ii) test characteristics and iii) statistical analysis. Foundation refers to the theoretical background of the test and was discussed in chapter 3.1.1. The second criterion of Athanasiou's framework, nonverbal test characteristics, include content review, modes of instructions, inclusion of practice items, response mode and use of timed items. These individual test aspects will be discussed in more detail below. Statistical properties related to cultural fairness is the empirical evaluation of cultural fairness and has to be assessed after the completion of the test battery.

Content review

According to Athanasiou (2000) content review refers to the detection of bias in the design of items. Common objects, such as animals, vehicles, houses, tools or clothes, often either do not look the same, or are more prevalent in one culture compared to other cultures. Furthermore, pictorial items may generally be unsuitable for individuals from cultures who are unaccustomed to representative drawings or pictures, as the response to those items might depend on past experiences with pictorial representations (Miller, 1973).

Therefore, test items depicting objects can be considered to have a tendency to be culturally biased.

Although non-pictorial tests, such as the Raven's progressive matrices and the Cattell Culture Fair Intelligence test include some items which show cultural bias (Nenty & Dinero, 1981; Powers, Jones & Barkan, 1986),

geometric patterns are less likely to be culturally biased than pictorial representations. For instance, Osuji (1982) showed that the performance of constructing complex geometric patterns is very similar for Canadian and Nigerian children. Consequently, pictorial representations were avoided and abstract geometrical patterns were used for the design of items in the CCIID.

Mode of instructions

Instructions communicate the requirements of a test. It is essential for the validity of the results that the participant fully understood the instructions of the task to ensure that the test assesses the cognitive ability the test was designed to assess and not the comprehension of the instructions (Braden & Athanasiou, 2005).

According to Jensen (1980) nonverbal tests require nonverbal instructions, which have to be pantomimed in order to communicate the instructions independently of acquired language skills. However, even the understanding of pantomimed instruction is not free from language skills (Oller, Kunok & Choe, 2001). Additionally, test instructions might not be understood fully if they are pantomimed only (Oller, Kim & Choe, 2001). Therefore, pantomimed instructions are deemed not sufficient to ensure that the participant would completely understand the tasks of the CCIID. In addition, standardized instructions, even if translated and back-translated correctly, can lead to a misinterpretation of the nature of the task when applied cross-culturally (Pena, 2007). Holding et al. (2004) showed that, depending on culture and social background, it is important, not only to change the instructions linguistically, but also in terms of teaching and corrections made during the examples. This led to the conclusion that also additional oral instructions, translated into the different languages of the participants, would not be sufficient to warrant that the tasks would be fully understood by all participants.

107

Subsequently, it was decided to use a two-tier approach to the instructions: instructions during the examples would be given in two forms (pantomimed and oral) and adjusted to the needs of the individual participant until the test assistant was sure that the participant had fully understood the task of the subtest. However, in the testing phase, only very limited and basic instructions would be used to remind the participants of the task. The specific instructions for each subtest will be discussed in section 3.2.

Inclusion of practice items

The inclusion of practice items was a central part of the instructions in order to ensure that the participant comprehended the nature of the task. Practice items were also found to reduce cultural bias (Jensen, 1980). Each subtest was, therefore, preceded by three examples to give the test assistant ample possibilities to explain the instructions for the subtest.

Response mode

Like the instructions, the response mode in a nonverbal test should also be given nonverbally (Athanasiou, 2000). The participant would use a computer pen or his/her finger to either choose options or to construct geometrical shapes. Consequently, the participant would complete the test without communicating with the test assistant. This form of response was, therefore, entirely nonverbal. Responses were recorded through the touch screen interface of the computer.

Timing of test items can lead to a misrepresentation of the ability assessed in the test, as it then becomes a measure of speed, rather than a measure of the ability required (Athanasiou, 2000). This can bias the results, particularly for individuals from cultures who are not familiar with focusing on speed as part of resolving a task (Harris, Reynolds & Koegel, 1996). Therefore, timing should be avoided in tests of cognitive abilities designed for cross-cultural use. None of the subtests of the computerized test battery were thus timed.

3.1.5. Analysis of nonverbal intelligence test design

Several nonverbal intelligence (sub-) tests were subsequently examined to find suitable designs for the subtests assessing the broad abilities 'fluid reasoning' and 'visual processing'. Firstly, all intelligence tests included in the introduction in chapter 1 of this thesis were analysed:

- The Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV: Perceptual Reasoning Index) (Wechsler, 2003)
- The Stanford-Binet Intelligence Scale-Fifth Edition (SB 5: Item books 1 and 2) (Roid, 2003)
- The Kaufman Assessment Battery for Children-Second Edition (K-ABC: Simultaneous and Planning Indices) (Kaufman & Kaufman, 2004)
- The Comprehensive Test of Nonverbal Intelligence (CTONI) (Hammill, Pearson & Wiederholt, 1997)
- The Snijders-Oomen Non-verbal Intelligence Test 5½-17 (SON-R) (Snijders, Tellegen & Laros, 1989)

- The Leiter International Performance Scale-Revised (Leiter-R) (Roid & Miller, 1997)

Additionally, the designs of the three following nonverbal intelligence tests were examined, as they are considered to be predecessors of nonverbal intelligence tests and are still used as a basis for the design of novel nonverbal intelligence tests (Flanagan & Kaufman, 2004).

- Raven's Standard Progressive Matrices (RSPM) (Raven, Raven & Court, 2003)
- Test of "g": Culture fair (CCF) (Cattell, 1949)
- Koh's block design test (Koh) (Hutt, 1932)

All subtests containing items measuring inductive reasoning or visual processing abilities were investigated for the following criteria:

i.   content used for the items (geometrical shapes)
ii. different types of subtests used to assess inductive reasoning or visual processing
iii. visual complexity of the design
iv. suitability for use on a touch-screen computer
v.   simplicity of the nonverbal instructions for the subtest.

Design choice for Inductive Reasoning subtest

For the selection of a subtest for inductive reasoning, the above criteria were employed:

i.   Drawing content

Most intelligence tests would use both pictorial as well as non-pictorial subtests to assess inductive reasoning (WISC-IV, SB5, K-ABC, CTONI, SON-R, Leiter-R). As discussed above, subtests containing pictorial representations, would not be suitable for the purpose of this test (see chapter 3.1.4. Cultural Fairness) and were, therefore, not considered for the development of the test battery. The following subtests for inductive reasoning were examined:

- WISC-IV: 'Matrix Reasoning'
- SB5: 'Object-Series/Matrices'
- K-ABC: 'Pattern reasoning'
- CTONI: 'Geometric Analogies', 'Geometric Categories', 'Geometric Sequences'
- SON-R: 'Analogies'
- Leiter-R: 'Classification',
- RSPM: all sets
- CCF: all subtests

ii. Types of subtests for inductive reasoning

According to Mayer (1994) psychometric tests for inductive reasoning can be clustered into the following groups: 'Classification problems', 'Analogy

Problems' and 'Series Completion Problems'. Nonverbal 'Classification problems' require the participant to find a common rule between two or more shapes or pictures. In nonverbal 'Analogy Problems', two sets of shapes or pictures of transformations are presented with one part of the second set missing. The participant has to apply the rule of transformation from the first set to the second set. Nonverbal 'Series Completion Problems' require the participant to complete a series of shapes or pictures that show a transformation.

Subtests using matrices (WISC-IV: 'Matrix Reasoning', SB5: 'Object-Series/Matrices', Raven's Standard Progressive Matrices, subtest 'Matrices' in Test of "g": Culture fair) are a combination of 'Analogy Problems' and 'Series Completion Problems'. All inductive reasoning tasks require the participant to follow a series of cognitive processes (encoding the presented problem, inducing a rule, applying this rule to a new case and responding by selecting the correct option) (Mayer, 1994; McGrew, 2005). Therefore, the different test types within this group of subtests were considered comparable as they should measure the same inductive reasoning abilities.

iii. Visual complexity of the design

Visual complexity refers to the amount of lines and detail in a drawing (Alario & Ferrand, 1999). Inductive reasoning tests using matrix-designs (WISC-IV: 'Matrix Reasoning', SB5: 'Object-Series/Matrices', RSPM, subtest 'Matrices' in CCF) are the visually most complex subtests as they show a matrix of 3 by 3 objects with one object missing. Additionally, 4 to 6 objects are presented as options to complete the matrix.

'Classification problems', 'Analogy Problems' and 'Series Completion Problems' contain considerably less visual complexity as they present less objects per item (between 5 and 8) :

112

'Classification problems' include the subtests 'Geometric Categories' (CTONI), 'Classification' (Leiter-R) and the subtest 'Classifications' in the CCF. This group of inductive reasoning tests showed between 5 and 8 objects depending on test.

The group 'Analogy Problems' included the subtests 'Geometric Analogies' (CTONI) and 'Analogies' (SON-R) which showed 7 or 8 objects.

The subtests 'Pattern Reasoning' (K-ABC), 'Geometric Sequences' (CTONI) and 'Series' (CCF) were 'Series Completion Problems' and presented 8 objects.

The number of objects presented in each item will influence the visual complexity of the item (Alario & Ferrand, 1999; Snodgrass & Vanderwart, 1980). Stimuli with a higher visual complexity require higher visual processing abilities than stimuli with a lower visual complexity (Harrison & Stiles, 2008; Mondloch, Geldart, Maurer & de Schonen, 2003).

A cognitive test battery designed for individuals with ID needed to include the easiest possible items in order to classify a wide range of cognitive abilities. A generally low visual complexity for a subtest is, therefore, preferable to a high visual complexity as it increases the possibilities for the design of easy items. As the matrix design used the highest number of objects (13-15), this test design has a higher degree of visual complexity and, therefore, would require more visual processing abilities than other designs for inductive reasoning subtests. Consequently, the matrix design was eliminated as a possible design for an inductive reasoning subtest.

iv. Suitability for use on a touch screen computer

The design of all three subtest types ('Analogy Problems', 'Classification Problems' and 'Series Completion Problems') was suitable for use on a touch screen computer.

113

v. Simplicity of nonverbal instructions

Considering the instructions for all three subtest groups, it became clear that instructions for 'Classification Problem' tests and 'Series Completion Problem' tests could be given nonverbally, by pointing and using confirmatory signs (nodding and smiling). 'Analogy problems', however, would need the explanation that the change happening to a first set of shapes has to be transformed to a second set of shapes. This would be difficult to describe in a nonverbal manner. Therefore, it was decided to use a 'Classification Problem' test and 'Series Completion Problem' test for the design of the CCIID.

Design choice for visual processing abilities

i. Drawing content

All visual processing subtests except 'Conceptual Thinking' (K-ABC) are non-pictorial and were, therefore, appropriate for the development of a culture fair test (see 1.4. Cultural Fairness). 'Conceptual Thinking' was not further considered for the design of the computerized test battery.
The following subtests were examined:

WISC-IV: 'Block Design'
SB5: 'Form Board'
K-ABC: 'Triangles',
SON-R: 'Mosaics', 'Patterns'
Leiter-R: 'Matching', 'Paper Folding', 'Figure Rotation'
Koh's Block Design Test

ii. Groups of subtests for measuring visuo-spatial abilities

The majority of the tests are similar in their concept: 'Block Design' (WISC-IV), 'Form Board' (SB5), 'Triangles' (K-ABC), 'Mosaics' (SON-R), 'Patterns' (SON-R), and Koh's Block Design Test all require the participant to replicate an abstract pattern, or to find an identical design ('Matching' – Leiter-R). This type of test is named 'Visualisation' task and commonly used in intelligence tests, as it is highly correlated to general intelligence scores (Cattell, 1971; Snow, Kyllonen & Marshalek, 1984; Royer, Gilmore & Gruhn, 1984). Consequently, it was decided to include a visualisation task in the development of the new test battery.

The subtests 'Paper Folding' and 'Figure Rotation' (both included in the Leiter-R) assess mental rotation. A mental rotation subtest is part of the wider cognitive test battery, and was, therefore, not considered as a design option for the CCIID.

iii. Visual complexity

'Visualisation' tasks differ in number of blocks used to replicate an abstract pattern. For instance, 'Block Design' (WISC-IV), uses 4 to 9 blocks depending on age and ability (Wechsler, 2003), while Koh's Block Design Test uses 4 to 16 blocks depending on age and ability (Hutt, 1932). Other tests like the SON-R keep the number of blocks consistent at 9 blocks. A 3 x 3 block design was chosen to create a maximum degree of difficulty. The number of blocks was kept constant to avoid confusing the participants with different layouts.

iv. Suitability for use on a touch screen computer

'Visualisation' tests can be administered in a two-dimensional form (pieces to replicate the abstract pattern are cards) or a three-dimensional form (pieces are blocks). A two-dimensional form would be suitable for use on a touch screen computer.

v. Simplicity of nonverbal instructions

'Visualisation' tests have been developed for deaf and hearing- impaired children using nonverbal instructions supported by examples (Snijders, Tellegen & Laros, 1989). A similar approach to instructions was employed in de development of a 'Visualisation' subtest for the CCIID to ensure that the instructions were clear to the participants with ID.

3.1.6. Interim conclusion

The development of the CCIID was based on five aspects: the Cattell-Horn-Carroll theory of cognitive abilities as a theoretical foundation for the test battery, the evaluation of the pilot project and studies 1 to 3, the target group (individuals with ID), cultural fairness and an examination of different designs of existing nonverbal IQ subtests.

The analysis of these aspects resulted in three different subtests: two subtests for inductive reasoning and one subtest for visual processing abilities, which together represent the two broad cognitive abilities 'fluid reasoning' and 'visual processing' in the Cattell-Horn-Carroll theory of cognitive abilities. Tests for these two broad cognitive abilities were selected as the results of the initial studies had indicated that these abilities were the best predictors for physical performance of athletes with ID.

Furthermore, the floor effect in the scores of the SON-R intelligence test used in the initial studies had shown the need for an instrument with additional, easy-level items to discriminate between lower degrees of cognitive abilities. As a result, out of wide range of existing types of subtests, those were chosen that would offer the possibility to develop easy-level items.

In addition, considering the target groups of individuals with ID, it was important to select tests with instructions that would be simple and easy to understand. Other subtest types with more complex designs, such as matrix designs, or subtest types with fairly complicated instructions, such as 'Analogy problems', were thus excluded from further consideration for the new test battery.

The test battery was developed for individuals with ID from different cultural backgrounds. Athanasiou's theoretical framework for cultural fairness (2002) was used for the development to keep cross-cultural bias of the CCIID to a minimum. Based on the framework, subtests for verbal abilities and general knowledge, as well as tests using pictorial representations were thus excluded. Furthermore, appropriate forms for instructions and practice items were developed to ensure cultural fairness of the test battery.

These considerations led to the selection of three types of subtests: A 'Classification Problem' and a 'Series Completion Problem' to assess inductive reasoning abilities, and a 'Visualisation' task to measure visual processing abilities. The development and design of these three subtests will be explored further in the following chapters.

3.2. Design of subtests

The following chapters will first specify shared aspects of the subtests (programming, examples, instructions, parallel versions, adaptive testing and scoring) and then report the research each subtest is based on (subtest

composition, the theory of difficulty, subtest specific instructions and examples).

3.2.1. Shared aspects of all subtests

Programming

All subtests were designed using The Gimp 2.2. software by the author of this thesis and programmed specifically for this purpose in C++ using OpenGL to implement the graphics layer, by Dr. Stephan Bandelow, Loughborough University. The initial interface is used to choose the appropriate subtest and to register participant details. It provided space for participant identification, age, sex and number of session (for retesting). Further buttons showed the different subtests with a drop down menu to select examples and subtests for assessment. Minor modifications were programmed by the author of the thesis.

Examples

Each subtest was preceded by three examples which would be used to support the instructions for the subtest. The examples started with an item of the easiest level of difficulty. Example items two and three increased in difficulty in order to prepare the participant for the change in level of difficulty and to explain that the same instructions should be applied to a more difficult level when items looked slightly different or more complicated. Although the more difficult examples might be challenging for some participants, the examples gave them the opportunity to understand how the instructions would apply to the more complex looking items.

Mistakes made during the example items would be pointed out. The participant would then be given the possibility to correct the mistake. If the participant failed to present the correct solution of the example item, the test

assistant would explain in more detail to the participant how to complete the example correctly. The purpose of the practice items was twofold: Firstly, they would be used to communicate the instructions and to familiarize the participant with the subtest and the use of the equipment (touch screen) for that subtest. Secondly, they would provide the test assistant with a control mechanism to monitor whether the participant understood the instructions correctly. Examples were completed as described for each subtest in the following chapters. All examples can be found in appendix C.

Instructions

As discussed in chapter 3.1.4, the diversity in ability and culture of the target population made it necessary to allow deviation from strictly standardized instructions during the examples (but not during the assessment). Although each subtest had its own basic instructions, which would be used for the testing phase, during the examples these could be complemented by the test assistant according to the needs of the individual participant. Instructions were adapted for each participant individually to their cognitive abilities and native language. Additionally, for the oral instructions during the examples, sports coaches could be included into the delivery of the instructions as they would speak the native language of the participant and be familiar with the ability of the participants to understand instructions, which are an important part of sports coaching. Throughout the examples, the coach could translate and support the instructions given by the test administrator. However, after the examples, when the test administrator was confident that the participant had understood the instructions, the coach would be thanked for their help, and politely, but firmly, be asked to not further interfere with the testing. Depending on the ability of the participant to understand the instructions and to score the example items correctly, one or more instruction modes were used (i.e. pantomimed and spoken - if necessary translated - instructions). If the test assistant was uncertain whether the participant understood the

instructions correctly, examples could be repeated until the test instructor was confident that the participant either understood the task of the subtest, or was not able to follow the instructions. When it was obvious that the participant was not able to complete the examples, the assessment was discontinued and the reason for discontinuation would be noted on a test report form. Instructions to all subtests can be found in appendix D.

Parallel versions

Based on the test construction of the SON-R 5 ½ - 17 intelligence test (Snijders, Tellegen & Laros, 1989) three parallel versions were developed for the subtests 'Odd One Out' and 'Series'. These three versions were administered in an adaptive testing procedure (see below) to limit the impact of correctly guessed answers on the result. The subtest 'Jigsaw' does not provide multiple choice options, which could be answered correctly by chance. Therefore, only one version was developed to assess the level of cognitive abilities.

Adaptive test procedure

An adaptive test procedure was chosen for the subtests 'Odd One Out' and 'Series' for several reasons:

1) The administration of parallel versions of the subtests reduces the impact of correctly scored items by chance. The subtests 'Odd One Out' and 'Series' were designed as multiple choice tests and, therefore, allowed the participant to select a random option. However, an adaptive test procedure reduces the number of items that could be guessed as items are only administered until the participant had scored two incorrectly.

2) Furthermore, the adaptive test procedure reduces the number of items that the participant experiences as too easy and therefore increases the

motivation of the participant to continue the assessment (Rust & Golombok, 1989).

3) In addition, an adaptive test procedure reduces administration time as the starting point for the second and third version depends on the participant's ability and is just below the level of difficulty the participant reached in the previous version.

Scoring

For the adaptive testing procedure of the subtests 'Series' and 'Odd One Out' participants started with version A and proceeded until they scored two items incorrectly, or until the version was finished. After two incorrect items, the version would be stopped regardless of having been scored incorrectly consecutively, or with correct items in between. After stopping the first version (either after two incorrect items or after the last item), the participant would continue at one level of difficulty below the number of correctly scored items. Therefore, the calculation for the starting point of the next version is: (number of correct items) minus 1. This scoring is similar to the SON-R (Snijders, Tellegen & Laros, 1989). The same procedure was applied to version two and three of both subtests. Items below the starting item in parallel version two and three were automatically scored as correct.

The subtest 'Jigsaw' was stopped after two incorrectly scored items (again, regardless of having been scored incorrectly consecutively, or with correct items in between) without proceeding to another version.

3.2.2. 'Classification Problem' - 'Odd One Out'

Subtest composition

Nonverbal 'Classification Problems' are a typical example of inductive reasoning and require the participant to find a common rule between two or more shapes or pictures. The 'Odd-one-out' subtest is a standard 'Classification Problem' based on concept- learning (Mayer, 1994). The

participant has to compare and contrast the different features in order to find the concept-relevant detail(s) that lead to the inference of a common rule which then can be applied to all but one shape.

The reason for selecting this particular type of 'Classification Problem' was the simplicity of its instructions, which are shorter and easier to communicate than in other types of 'Classification Problems' tests. Each item of the subtest 'Odd One Out' contained six geometrical shapes displayed in two rows on the computer screen. Five of these shapes shared one or more common features, which the sixth item would not share. All shapes consisted of black lines on a white squared background, which were presented in two rows of three on a black screen. The subtest's three parallel versions each included 15 items designed to represent increasing levels of difficulty. The subtest was preceded by three examples. The 'Odd One Out' shape was placed in a randomized position.

Theory of difficulty

The theory of difficulty was based on three factors, which influenced the level of difficulty for this subtest:

- Visual complexity
- Discriminability of differences
- Number of distractors

A. Visual complexity

As discussed in chapter 3.1.5, the level of visual complexity affects the degree of required visual processing abilities (Harrison & Stiles, 2008; Mondloch, Geldart, Maurer & de Schonen, 2003). Therefore, items were designed to increase in visual complexity. Each version started with simple shapes consisting of only one single line (circle, rectangle, triangle) and

showed an increase in the number of lines and details in items designed to represent a higher degree of difficulty.

B. Discriminability of differences

Shapes or features within shapes can have different degrees of discriminability, i.e. the difference between target ('Odd-One-Out') and non-target shapes can be quite obvious, such as differences in size, colour (black or white) and shape, which are easy to detect, or subtle, such as differences in position (e.g. upside down) or of the relation of objects to each other, which are more difficult to notice. This is reported to influence the difficulty of an item (Duncan & Humphreys, 1989). Consequently, the discriminability of shapes or features within shapes was used to vary the degree of difficulty of items.

C. Number of distractors

Distractors are features that are not relevant for the inference of the common rule that leads to the identification of the 'Odd One Out'. Their function is to degrade the processing ability. Particularly in complex serial processing, the number of visual distractors influences the cognitive processing ability (Treisman, 1991; Treisman & Gormican, 1988). Therefore, it was decided to use an increasing number of distractors for items on a more difficult level of the test battery.

Based on these three factors (visual complexity, discriminability of differences and number of distractors), the original form of the subtest 'Odd One Out' included 15 items (each containing six pictures: five plus the 'Odd One Out') using increasing levels of difficulty:

Level 1: All six geometrical pictures are based on one single shape: 5 pictures are identical – the 'Odd One Out' is different in colour, shape and size

Level 2: All six geometrical pictures are based on one single shape: 5 pictures are identical – the 'Odd One Out' is different in shape and size

Level 3: All six geometrical pictures are based on one single shape: 5 pictures are identical – the 'Odd One Out' is different in size and slightly different in shape (e.g. oval instead of circle)

Level 4: All six geometrical pictures are based on one single shape: 5 pictures are identical – the 'Odd One Out' is different in size or shape

Level 5: All six geometrical pictures are based on one single shape: 5 pictures are identical – the 'Odd One Out' is slightly different in shape

Level 6: All six geometrical pictures are based on two single shapes: One of those is identical in all six pictures (distractor), the second shape is identical in 5 pictures – the second shape of the 'Odd One Out' is different

Level 7: All six geometrical pictures are based on two single shapes: One of those is identical in all six pictures (distractor), the second shape is identical in 5 pictures – the second shape of the 'Odd One Out' is slightly different

Level 8: All six geometrical pictures are based on two single shapes: One of those is clearly varying in all pictures (distractor), the second is identical in 5 pictures – but slightly different in the 'Odd One Out'

Level 9: All six geometrical pictures are based on two single shapes: One varies but is not different in concept in all pictures (e.g. all triangles but with different angles = distractors), the second shape varies in position in 5 pictures and varies in concept in the 'Odd One Out' (e.g. parallel lines with different distances vs. crossing line in the 'Odd One Out')

Level 10: All six geometrical pictures are based on two single shapes: One varies but is not different in concept in all pictures (e.g. all triangles

124

but with different angles = distractors), the second shape is different in all six pictures but also varies in concept in the 'Odd One Out' (e.g. third shape is touching vs. not-touching in the 'Odd One Out')

Level 11: All six geometrical pictures are based on three single figures: Two features within the figures vary (e.g. colour and shape = distractors) – 'Odd One Out' has an additional feature that is clearly different from the other figures (e.g. size)

Level 12: All six geometrical pictures are based on three single figures: Two features within the figures vary (e.g. colour and shape = distractors) – 'Odd One Out' has an additional feature that is slightly different (e.g. up-side-down)

Level 13: All six geometrical pictures are based on three single figures: Three features within the figures vary (e.g. colour, size and shape = distractors) – 'Odd One Out' has an additional feature that is clearly different (number of coloured features)

Level 14: All six geometrical pictures are based on three single figures: Three features within the figures vary (e.g. colour, size and shape = distractors) – 'Odd One Out' has an additional feature that is slightly different (touching vs. not-touching)

Level 15: Several shapes which share common and different features (distractors) – 'Odd One Out' does not share a specific common characteristic.

Examples

The subtest 'Odd One Out' was preceded by three examples, which increased in difficulty. The first example showed five identical shapes and one very different shape, the 'Odd One Out' (five small black squares and one large white oval was the 'Odd One Out'). This example was designed to show

125

that five shapes belong together, but that one is different. This first example was equivalent to a level 1 item. The second example presented five white, large shapes that differ in form (circle, oval, square, triangle, octagon), but that were constant in size and colour. The 'Odd One Out' was a small, black rectangle. This example was constructed to demonstrate that while all shapes can be slightly different, the five objects share more features with each other than with the 'Odd One Out'. The example was equivalent to a level 8 item. The third example presented six different, black shapes, similar in size with five of them containing a white stripe and one of them (square) containing two white triangles. Again, this example was designed to confirm the notion that all shapes can be different, but that the 'Odd One Out' is even more different (white triangles instead of stripes). The example was equivalent to a level 9 item. Although example two and three were likely to be challenging for some participants, they provided an illustration of the increasing levels of difficulty, and showed different features of the test (that the different shapes in a picture can vary, but that the 'Odd One Out' is most different or varies in a different manner). Therefore, it was important to include examples with increasing levels of difficulty to familiarise the participant with the subtest. The examples can be found in appendix C.

Instructions

Like for the other subtests of the CCIID, instructions were provided in two forms: First, basic pantomimed gestures were used to explain the instructions supported by examples. Second, simple instructional phrases were – if necessary - translated into the native language of the participant by the test instructor or by English speaking coaches (see chapter 3.2.1.).

For the pantomimed gestures the test administrator would first encircle the five pictures excluding the 'Odd One Out' and then point at the 'Odd One Out' to indicate that this object does not belong with the others. The participant

would select the 'Odd One Out' by pointing to it with his finger (but without touching the computer screen in case of screens reacting to finger touch instead of computer pens). Only when the correct option was chosen, would the test administrator encourage the participant to touch the screen or hand the computer pen over to the participant. Then, the participant would be encouraged to point at the (correct) option again, using the pen. This was important because if the participant used the computer pen to point at an incorrect option, the test administrator would not have time to correct the mistake. For computer screen interfaces working by finger touch, the test administrator would put his hand in front of the screen to stop the participant from making a wrong choice and to correct the mistake. The pantomimed instructions remained the same for all three examples.

The verbal instructions for the first example were: *Here you see six shapes, five of them belong together, and one of them is different. Which one is different?* If the native language of the participant was not English, and the test administrator was not fluent in the participant's native language, the test administrator asked the coach to translate the verbal instructions. The verbal instructions for the second example would emphasise that all shapes are now different: *Now the shapes are all different but one is very different. Which one is it?* For the third example the test administrator would point out that although all shapes were very different, now they would share a common feature: *Now all shapes are very different, but they have one thing in common, but one shape does not. Which one is that? Which one is the 'Odd One Out'?* During the examples, the coach was allowed to help to encourage the participant and was given the possibility to explain the instructions in the native language of the participant.

Depending on the response of the participant the test administrator would react accordingly:

If the participant picked the right shape, the test administrator would say: *Well done!* and would move on the next example, or, after the last example, to the first test version.

If the participant would choose the incorrect shape, the test administrator would first encourage the participant to choose again, and if the second choice was incorrect again, point out the correct shape: *This one is different!*

If the participant did not react despite encouragement, the test administrator would point out the 'Odd One Out' shape to the participant. Again, the test administrator should encircle the five shapes excluding the 'Odd One Out' and then point to the 'Odd One Out' and say: *This one is different.* (if necessary translated by the coach).

If the participant pointed to two or more shapes the test administrator would say: *Only one is most different. Which one is it?* (if necessary translated by the coach).

If necessary, basic test instructions were used after the examples if necessary to remind the participant what to do (*'Which one is most different?'*). The test administrator would not provide corrections after the completion of the examples.

3.2.3. 'Series Completion Problem' - 'Series'

Subtest composition

The nonverbal subtest 'Series' was designed as another test to assess inductive reasoning abilities and belongs to the group of 'Series Completion Problems' (Mayer, 1994). The participant had to choose from four options, one that completes a series of four transforming shapes. 'Series' required the participant to find similarities and/or dissimilarities of shapes and geometric features within the relationships between objects (De Koning, Sijtsma & Hamers, 2003).

128

Each item of the subtest 'Series' contained two rows of four geometrical shapes (or a combination of shapes). The upper row showed the transformation of a shape during the first three pictures (from left to right) and a question mark in the place of the fourth shape. The lower row showed four options from which the participant had to choose the one option that should be in the place of the question mark in order to complete the series of transformations. The position of the correct option in the bottom row was randomized for all items. All shapes and combinations of shapes were drawn in black lines on white squares which were presented on a black screen (see appendix F). Again, the subtest consisted of three parallel versions, each including 16 items designed to represent increasing levels of difficulty. The subtest was preceded by three examples.

Theory of difficulty

The theory of difficulty was based on four factors influencing the level of difficulty:

- Visual complexity
- Type of transformation
- Number of simultaneous transformations
- Difficulty to detect the correct option

A. Visual complexity
As with the subtest 'Odd One Out', the degree of visual complexity influences the level of requirement for visual processing abilities to detect similarities and dissimilarities in the relationships between the shapes (Harrison & Stiles, 2008; Mondloch, Geldart, Maurer & de Schonen, 2003). Therefore, items were designed to increase in visual complexity. Each of the three versions started with simple shapes consisting of only one single line (circle, rectangle,

triangle) and, subsequently, had an increased number of lines and details to lead to a higher degree of difficulty.

B. Type of transformation

Mental rotations of shapes are considered to require more complex processing abilities than changes of size (Rösler, Heil, Bajric, Pauls & Henninghausen, 1995). Consequently, transformations based on mental rotation were considered to be more difficult than transformations based on size scaling.

C. Number of simultaneous transformations

The details of a shape can undergo one or more transformations simultaneously, e.g. a black corner in a square rotates while the scale of the square increases. An increase in the number of simultaneous visual tasks is regarded to increase the difficulty of the cognitive task (Boles & Law, 1998).

D. Difficulty to detect the correct option

The options were presented using different levels of difficulty to detect the correct option: One option was obviously incorrect and therefore was easy to detect as such. A second option was designed on an intermediate level, and a third option was devised to be difficult to rate as incorrect as it resembled the correct option quite closely. The degree of difficulty of the options was kept constant throughout the items.

Based on these four factors (visual complexity, type of transformation, number of simultaneous transformations, difficulty to detect the correct option), the original form of the subtest 'Series' included 16 items on increasing levels of difficulty:

Level 1:   All three pictures contain the same basic shape which is identical, there is no transformation

Level 2: The scale of a basic shape increases or decreases

Level 3: The shape contains one simple rotating feature

Level 4: One part of the shape in- or decreases in size while the rest remains on the same scale as a distractor

Level 5: The shape contains one rotating and one distrator feature

Level 6: Two shapes per picture: One remains constant (the distractor), one changes in scale

Level 7: Two shapes per picture: One remains constant (the distractor), one rotates

Level 8: Two shapes per picture: Both change in scale

Level 9: Two shapes per picture: One changes in scale, one rotates

Level 10: Two shapes per picture: Both rotate

Level 11: Three shapes per picture: One remains constant (the distractor), one changes in scale, one rotates

Level 12: Three shapes per picture: One remains constant (the distractor), two rotate

Level 13: Three shapes per picture: Two shapes change in scale, one rotates

Level 14: Three shapes per picture: One shape changes in scale, two rotate

Level 15: Three shapes per picture: All three shapes rotate

Level 16: Three shapes per picture: Four shapes, three rotate, one remains constant (the distractor)

Examples

For the design of the examples, it was decided that the three practice items would represent the first three levels of difficulty of the subtest, to introduce the participant to different concepts of transformations. The first example showed an item which was equivalent to difficulty level 1. This example was intended to convey the message that the shapes of the upper row formed a

131

meaningful line with one picture missing while the bottom row contained options which the participant has to choose from. The second example was based on difficulty level 2 and was designed to demonstrate the concept of transformation in the upper row of the display. The third example presented an item, which was equivalent to difficulty level 3 and was aimed to show, again, the concept of transformation in the upper row and to introduce another kind of transformation (rotation). Although example two and three were likely to be challenging for some participants, they provided an illustration of the increasing levels of difficulty, and showed different features in the test (mainly that the transformation can occur in different forms). Therefore, it was important to include examples with increasing levels of difficulty to familiarise the participant with the subtest. The examples for the subtest 'Series' can be found in appendix C.

Instructions

Basic pantomimed gestures were used to explain the task of the subtest supported by examples. The test administrator would first point at the upper row of shapes and then at the question mark at the end of the row. Next, the test administrator would point over the options and go back to the question mark and, then, give the participant time to choose an option. The participant would select the option by pointing to it with his finger (see chapter 3.2.1.). The pantomimed instructions remained the same for all three examples.

The verbal instructions for the first example were: *In the row up here, you see three shapes and a question mark* (point). *Which one of these shapes* (point to the options) *belongs in the place of the question mark?*

The verbal instructions for examples two and three were: *Now these shapes are changing* (point to the upper row). (Point to first shape) *This changes into that* (point to second shape) *and that changes further into that*

132

(point to third shape) *and that changes further again* (point to question mark). *How would it look like out of those?* (point to options).

If the answer was correct, the test administrator would say: *Well done!* The test administrator would move on to the next example or, after the last example, to the first test version.

If the answer was incorrect, the test administrator would first encourage the participant to choose again, and, if the second choice was incorrect again, point out the correct shape: *Look, it is this one.* (Points to the first shape and explain again:) *This changes into that* (points to second shape) *and that changes further into that* (points to third shape) *and that changes further into this one* (points to the correct option).

If the participant did not react despite encouragement, the test administrator would point out the correct option to the participant.

Test instructions were used after the examples, if deemed necessary. However, instructions would only be given to remind the participant what to do. The test administrator would not correct the participant after the completion of the examples.

3.2.4. 'Visualisation task' – 'Jigsaw'

Subtest composition

The visualisation subtest 'Jigsaw' was designed to measure visual processing abilities (see chapter 3.2.4). Each item showed a geometric picture composed of 3x3 uni-coloured and/or patterned squares on the left side of the screen. On the right side of the screen, an empty frame of the same size as the 3x3 pattern was shown. On the bottom of the screen were blue and yellow uni-coloured and/or patterned squares to reconstruct the picture on the left into the frame on the right. The task was to move these pieces into the correct position to copy the picture on the left by first tapping the required piece at the bottom of the screen, and then, tapping the position in the frame on the right that would show this piece. Additionally, the screen showed a square

133

containing two turning arrows (↻), which the participant could tap in order to turn patterned pieces into the correct position. Tapping the arrows once would turn a piece 90°, tapping it twice 180°, three times 270° and four times would turn it back into its original position. There was no limitation set on how often each piece could be turned. As soon as the frame was filled with nine squares, an upwards pointing arrow (↑) would appear between the original picture and the filled frame to indicate that the participant could move on to the next item by tapping this arrow. Each item would only show those pieces necessary to re-construct the picture. The participant had 120 seconds to complete the item. If he/she exceeded the time limit, the item would be scored as 'incorrect'.

The geometric picture and all pieces were blue and/or yellow on a black screen. The frame was outlined in orange. The arrows were green on a white, squared background (see appendix E). The subtest was preceded by three examples.

Theory of difficulty

The theory of difficulty was based on four factors influencing the level of difficulty:

- Symmetry
- Different types of pieces
- Rotation
- Gestalt theory

A. Symmetry

A symmetric picture is easier to re-construct than an asymmetric display (Schorr, Bower & Kiernan, 1982). Therefore, it was decided that a symmetric picture would only be used as the first and second item. Subsequent items were designed to be asymmetrical to increase the degree of difficulty.

134

B. Different types of pieces

The subtest contained two types of pieces, uni-coloured squares (blue and yellow) and four different, blue-yellow patterned squares. The use of patterned pieces resulted in a more visually complex picture than the use of uni-coloured pieces, and different kinds of patterned pieces created more complexity. The use of patterned pieces, therefore, increased the level of detail of the visual information (Marois & Ivanoff, 2005).

C. Rotation

Patterned pieces were also used in non-rotated or rotated positions. Mental rotations required more cognitive abilities than the positioning alone without rotation. Therefore, rotation was used as a factor influencing the degree of difficulty.

D. Gestalt theory

According to Gestalt theory global patterns with a higher degree of element integration influence the difficulty of visualisation tasks (Schorr, Bower & Kiernan, 1982). However, the exact interaction between pattern integration and degree of difficulty has not yet been established. Therefore, global patterns were generally avoided.

The three factors symmetry, different types of pieces and rotation determined the theory of difficulty of the subtest 'Jigsaw' which included 11 items on increasing levels of difficulty:

Level 1: Vertically and horizontally symmetric, using uni-coloured pieces

Level 2: Vertically or horizontally symmetric, using uni-coloured pieces

Level 3: Asymmetric, using uni-coloured pieces

Level 4: Asymmetric, using uni-coloured and 2 patterned pieces of the same type which have to be rotated

Level 5: Asymmetric, using 5 uni-coloured and 4 patterned pieces of the same type

Level 6: Asymmetric, using 3 uni-coloured and 6 patterned pieces of 2 different types

Level 7: Asymmetric, using 1 uni-coloured piece and 8 patterned pieces of 2 different types

Level 8: Asymmetric, using 1 uni-coloured piece and 8 patterned pieces of 3 different types, 5 rotations

Level 9: Asymmetric, using 1 uni-coloured piece and 8 patterned pieces of 3 different types, 6 rotations

Level 10: Asymmetric, using 1 uni-coloured piece and 8 patterned pieces of 4 different types, 7 rotations

Level 11: Asymmetric, using 1 uni-coloured piece and 8 patterned pieces of 4 different types, 8 rotations

Examples

Three examples were completed to familiarise the participant with the subtest. The first example included uni-coloured and patterned pieces (rotated). It would be completed by the test assistant to demonstrate the different features of the subtest and to show:

- How pieces should be moved into the frame

- How pieces should be rotated
- How mistakes could be corrected
- How the participant would move to the next item

The examiner would slowly complete the first example while making certain that the participant paid attention. Then, the test assistant would encourage the participant to start on the second example and – in case of a touch-screen operated with a computer pen - hand over the pen to the participant. The second example contained a simple pattern (cross) based on uni-coloured pieces to familiarise the participant with the movement of the pieces while putting little cognitive demands on pattern re-construction. The test assistant would support the participant, using verbal or pantomimed instructions (verbal instructions would be translated by the coach, if necessary), further demonstrations or manual aid to move the pen. The picture of the third example contained four patterned pieces. The participant would need to rotate three patterned pieces to copy the picture correctly. Again, the test assistant would support the participant, if required. Although example three was likely to be challenging for some participants, it provided an illustration of the increasing level of difficulty and showed different features in the test (rotation of pieces). Therefore, it was important to include a more difficult example to familiarise the participant with the subtest. The examples can be found in appendix C.

Instructions

While completing the first example the test administrator would explain the different features of the subtest when appropriate. First, the test administrator would point at the geometric picture on the left and say: *The task here is to copy this picture into that frame* (pointing at empty frame) *using these pieces* (pointing at pieces). *You move them like this* (the examiner starts filling the

frame, makes a mistake on purpose in upper row but continues to fill in pieces into the first two rows from the top).

After 2nd row: *You can only change the position of the last piece* (test assistant demonstrates that only the piece that had just been moved into the frame can be moved again, not the others). *You can see, I made a mistake here. I can't change these pieces anymore, so I just put new ones over it* (test assistant corrects mistake). *If you want to change a piece, just put a new piece over it*

Before starting on the third row the test assistant points out the patterned pieces in the picture on the left and points at the patterned piece on the bottom of the screen: *You also need patterned pieces for some pictures* (test assistant points at picture) *like this one which you can turn* (points at piece at the bottom of screen). *When you tick that arrow* (test assistant points out arrow) *you can turn the patterned piece into the right position* (test assistant moves pieces into the right position and shows rotation)

When the frame is full, this arrow appears (points at arrow between the two pictures). *It does not mean that the picture is the same as this one (point) it just means the frame is full. Just check if the two pictures are the same if this is so, tick this arrow and the next picture will appear.*

After the first example the test assistant would say: *Now you try, please.* Following the first example, the test assistant would hand the pen over to the participant to continue with example two and three. Instructions were repeated and supported by the sports coach when necessary. Corrections were made if needed.

After the examples, instructions were given to remind the participant of the task (*Please copy this picture into that frame*) and to reiterate how to correct mistakes if the participant tried to change pieces (*Use a new piece*). No corrections were given after the example items.

3.3. Discussion

The CCIID was developed as the initial studies indicated that other cognitive test batteries were not suitable for the assessment of individuals with ID from diverse cultural backgrounds.

Firstly, the results of the initial studies had shown that even a nonverbal intelligence test is not sufficiently cultural fair if pictorial representations are included. Most intelligence tests, verbal and nonverbal, include pictorial representations. Few intelligence tests, such as the Raven's Coloured Progressive Matrices (RCPM) (Raven, Raven & Court, 1998), the Raven's Standard Progressive Matrices (RSPM) (Raven, Raven & Court, 2003) or the test of "g": the Catell Culture fair test (CCF) (Cattell, 1949), are based on geometrical and abstract shapes alone. However, these tests were not suitable for the research project as they were either developed for a very limited age range (RCPM: 5 to 11 years of age) or for adults only (CCF, RSPM).

Secondly, the nonverbal SON-R intelligence test, which was selected after a review of several verbal and nonverbal intelligence tests, had floor – effects in all subtests. Additional items on different easy levels were needed to ensure that the test battery would discriminate in the lower range of cognitive abilities.

Consequently, the Computerized Cognitive test battery for Individuals with ID (CCIID) was developed. All subtests were designed using abstract and geometrical shapes in order to minimize cultural bias and included an adequate number of items to discriminate in the lower range of cognitive abilities.

In order to endorse the content validity of the CCIID, test development was based on the Cattell-Horn-Carroll theory of cognitive abilities (CHC), and the designs for the subtests were chosen on basis of an analysis of the most commonly used nonverbal intelligence tests (see chapter 3.1.5). Although the

CCIID in itself measures only two broad cognitive abilities included in the CHC, fluid reasoning and visual processing, it is designed to be combined with a wider cognitive test battery, which also assesses short-term memory, long-term storage and retrieval, processing speed and decision speed/reaction time, but which is not part of this thesis.

A further advantage of the CCIID is its computerized form, which improves the administration as well as the access to the results. Mistakes made in the scoring process in pen and paper intelligence tests, such as the SB5, the SON-R or the Leiter-R, where scores have to be calculated during the assessment to determine the starting point for the next test, cannot happen in a computerized form, and, therefore, the scores are more reliable. In addition, scores can be calculated instantly by the software and do not have to be transferred from the paper into a computer, or be calculated by hand, which can lead to additional mistakes. The computerized form also reduces administration time, as the changes between subtests are done quicker on the computer than when using different booklets.

In sum, the CCIID was developed specifically for the assessment of individuals with ID from diverse cultural backgrounds. The theory-based structure supports content validity as well as cultural fairness of the CCIID.

The following sections will examine the psychometric properties of the CCIID to establish item difficulty, reliability and validity using modern and classical test theories.

3.4. Assessment of psychometric properties of the CCIID

Following the construction of the test battery, its psychometric properties were assessed for individuals with ID using modern and classical test theories. As discussed in chapter 1, the British term 'learning disability' is used interchangeably in British scientific literature with the term 'intellectual disability' (Cooper, Melville, Morrison, 2004; Cornwell, 2004). For the benefit of continuity, this thesis will use the term 'intellectual disability'. Any effects of the discrepancy of definitions between British and DSM-V standards will be discussed in section 3.4.4.

The assessment process of psychometric properties consisted of the following steps: Firstly, the items included in the CCIID were investigated using latent trait models and proportion of correctly scored items. Based on the results, several items that showed insufficient quality were re-designed. Subsequently, reliability and validity of the CCIID were examined and evaluated.

3.4.1. Item analysis

The item analysis was based on a combination of latent trait models and proportion of correctly scored items per parallel version for the subtests 'Series' and 'Odd One Out'. Latent trait models were used to determine several characteristics (=latent traits) of the items, such as difficulty and discriminatory ability. In addition to the latent trait model analysis, the proportion of correctly scored items was examined to analyse the difficulty level of each item, as it was not possible to obtain a sufficient large data set within the time frame for this thesis. Therefore, the exact misfit of particular items could not be determined based on the latent trait models, and the

141

results might indicate a lower number of misfitting items than a larger sample would show (Hambleton & Rovinelli, 1973; Reise & Yu, 1990).

For the subtest 'Jigsaw', item difficulty was established using proportions of correctly scored items per parallel version only, as the sample size was too small for the use of latent trait models. The following chapter will present the investigation of the test battery on an item level.

Methods

Participants

For the examination of the subtests 'Series' and 'Odd One Out', 69 participants, (29 male, 40 female) between 15 and 44 years of age, with different levels of schooling, were assessed. The study included 28 service users of a charity which provides education and employment for individuals with different levels of intellectual disability. The assessment also included 25 students from a local college (18 A-level students, 7 level 1 or 2 students receiving learning support) and 22 university students to assess the higher ability levels and increase variance within the sample. For the investigation of the subtest 'Jigsaw', 15 participants (7 male, 8 female) between 12 and 28 years of age were tested, 10 attended a special needs school for children and teenagers with different levels of ID and 5 students were from a local collage (A-level students).

Participants and -in case of power of attorney- their parents or carers had been given information about the purpose of the study and gave informed consent prior to the study. The studies included participants with different causes for intellectual disability as specific causes of the disability can only be established in a minority of cases (Kaski, 2009).

Instruments

The subtests of the CCIID were administered as described in section 3.2. All participants were tested using the non-adaptive form of the CCIID testing procedure, i.e. each participant completed each item in all three parallel versions of the subtests 'Series' and 'Odd One Out' and all items of the subtest 'Jigsaw', regardless of incorrectly scored items.

Procedure and environment

To avoid attention related problems due to long administration times, the subtests 'Series' and 'Odd One Out' were combined and administered separately from the subtest 'Jigsaw'. Due to the non-adaptive testing procedure the testing took about 20 minutes for each of the subtests 'Series' and 'Odd One Out' and 30 minutes for the subtest 'Jigsaw'. 'Series' and 'Odd One Out' were administered on two occasions at different Linkage locations (in the meeting room at the main office, and in the common rooms of supported living training houses), on two occasions in a waiting room and on four occasions in a meeting room. All testing rooms were quiet, large rooms. At one testing site, several participants were tested simultaneously in one room with sufficient space to not disturb each other's performance. The subtests were administered by trained test assistants (the author and final year psychology students) and on all occasions participants were given sufficient time to familiarize themselves with the test environment.

Statistical Analysis

The items of all subtests were scored dichotomously (as correct or incorrect). Scores on the subtests 'Series' and 'Odd One Out' were used for a latent trait model analysis, and for all subtests, the proportions of correctly scored items were computed. Due to low participant numbers for the subtest 'Jigsaw' no latent trait model analysis was performed for this subtest.

Descriptive statistics were calculated and assumptions and model-data fit were assessed. Two- and three-parameter models were computed and their fit was compared using a likelihood ratio test. Subsequently, difficulty and discrimination coefficients per item were obtained and in one case, for the better fitting 3-parameter model, also the guessing parameter.

For the evaluation of items, difficulty and discrimination parameters were assessed, and proportions of correctly scored items were examined. The results of the latent trait models established item difficulty using the difficulty coefficient with lower values representing easier items and higher values representing more difficult items. Discrimination properties of the items were investigated using the discrimination coefficient with values between 0.8 and 2.5 to be considered as 'good values' (De Ayala, 2009). For version 3 in the subtest 'Series', the guessing parameter was also examined, which should be lower than the chance value of guessing an item correctly (for subtest 'Series' four options were offered and therefore the guessing parameter should be lower than .25). However, most weight in the evaluation of items was given to the proportions of correctly scored items, as an exact misfit of items could not be determined due to the small sample size for the latent trait model analysis. Therefore, discrimination and difficulty parameter values were only used as indicators, while the decision for alteration of items was mainly based on the proportions of correctly scored items and the results of the latent trait model analysis were only considered if they had 'extreme values' or if the proportions of correctly scored items were inconclusive. 'Extreme values' for difficulty parameter values, were values above 2.0 or -2.0, for discrimination parameter, values above 10.0 and for the guessing parameter, values above 0.3. The results of the latent trait models can be displayed in figures or graphs (item characteristic curves). However, a presentation of the results in figures enables a more precise interpretation of the parameters. Therefore, the results of the latent trait model analysis of this thesis are presented in figures.

An example of an item characteristic curve (for version 1 of the subtest 'Series') is provided in appendix F.

For all analyses, a level of significance of 0.05 was used and all analyses were conducted in R.

Results

The item analysis for the subtests 'Series' and 'Odd One Out' included 69 participants (29 male, 40 female) with an age range between 15 and 44 years, a mean age of 23.74 and a standard deviation (SD) of 6.42. For the analysis of the proportions of correctly scored items of the subtest 'Jigsaw', 15 participants were assessed (7 male, 8 female) between 12 and 28 years of age, with a mean age of 16.33 and a standard deviation of 4.00. Descriptive statistics and numbers of participants per version are displayed in table 28.

Table 28 Descriptive statistics for item analysis

		N	Minimum score	Maximum score	Mean	SD
Subtest	Version A	67	2	16	10.18	4.35
'Series'	Version B	67	0	16	9.81	4.25
	Version C	68	1	16	9.69	4.48
Subtest	Version A	66	1	14	9.99	3.53
'Odd One	Version B	64	1	13	9.00	3.30
Out'	Version C	65	2	13	9.37	3.22
Subtest 'Jigsaw'		15	0	11	5.13	3.56

Assumptions for latent trait model analysis are uni-dimensionality, local independence[1] and a sufficient large set of data (Hambleton, 1991). Uni-dimensionality was assumed, as uni-dimensionality was confirmed for similar nonverbal intelligence test (e.g. SON-R; Snijders, Tellegen & Laros, 1989). Uni-dimensionality could not be empirically tested, as the assessment of the subtests was split into two sets ('Series' and 'Odd One Out' were combined while 'Jigsaw was tested separately). Therefore, a factor analysis confirming uni-dimensionality empirically, was not possible. Based on the uni-dimensionality, local independence was assumed (Lord, 1980).

For the latent trait model analysis, the fit of two models for the data was computed, a two-parameter latent trait model and a Birnbaum three parameter model (De Ayala, 2009; Rizopoulos, 2006). A likelihood ratio test was computed between the fit parameters (Log likelihood (Log Lik), Akaike information criterion (AIC) and Bayesian information criterion (BIC)) of both models, which showed that there was no significant difference for model-data fit between the 2- and the 3-parameter model for all versions, except for version 3 in the subtest 'Series', where the three-parameter model was a better fit (see table 29).

Table 29 Likelihood Ratio Table for version 3 subtest 'Series'

Model	AIC	BIC	Log Lik	LRT	df	p-value
2-parameter model	1125.54	1197.03	-530.77			
3-parameter model	1125.38	1232.62	-514.69	32.16	16	.01

---

[1] Local independence implies that all manifest variables are independent random variables if the latent variable is controlled

In case of equal fit of models, De Ayala (2009) suggested to choose the models with less parameter. Consequently, two-parameter models were chosen for all versions in the subtests 'Series' and 'Odd One Out', except for version 3 of the subtest 'Series'.

For the evaluation of items, difficulty and discrimination parameters were assessed, and proportions of correctly scored items were examined. The results of the latent trait models established item difficulty using the difficulty coefficient with lower values representing easier items and higher values representing more difficult items. Discrimination properties of the items were investigated using the discrimination coefficient with values between 0.8 and 2.5 to be considered as 'good values' (De Ayala, 2009). For version 3 in the subtest 'Series', the guessing parameter was also examined, which should be lower than the chance value of guessing an item correctly (for subtest 'Series' four options were offered and therefore the guessing parameter should be lower than .25). However, most weight in the evaluation of items was given to the proportions of correctly scored items, as an exact misfit of items could not be determined due to the small sample size for the latent trait model analysis. Therefore, discrimination and difficulty parameter values were only used as indicators, while the decision for alteration of items was mainly based on the proportions of correctly scored items.

Table 30 displays difficulty and discrimination parameters and proportions of correctly scored items for version 1 of the subtest 'Series'.

Table 30 Subtest 'Series' version 1

Item number	Difficulty coefficient	Discrimination coefficient	Proportion correct
1	- 1.35	2.48	0.82
2	- 1.25	2.22	0.81
3	- 0.67	10.36	0.67
4	- 0.35	0.92	0.55
5	- 0.65	30.60	0.64
6	- 0.73	0.97	0.63
7	- 0.30	0.78	0.54
8	- 0.64	1.27	0.63
9	- 0.78	3.07	0.70
10	- 0.79	2.77	0.70
11	- 0.77	4.47	0.70
12	- 1.04	1.00	0.69
13	- 0.26	1.11	0.54
14	- 0.96	0.47	0.60
15	- 0.53	1.16	0.60
16	0.84	0.72	0.36

The difficulty coefficients and proportion correctly scored items for version 1 of the subtest 'Series' indicated that the items 3 to 15 did not gradually increase in difficulty. Additionally, the discrimination coefficients of item 3 and 5 showed extreme values (see section 3.4.1), as they exceeded the range between 0.8 and 2.5 considerably (De Ayala, 2009).

Table 31 displays difficulty and discrimination parameters and proportion of correctly scored items for version 2 of the subtest 'Series'.

Table 31 Subtest 'Series' version 2

Item number	Difficulty coefficient	Discrimination coefficient	Proportion correct
1	- 1.82	2.75	0.94
2	- 0.23	1.14	0.55
3	- 0.77	1.67	0.70
4	- 0.54	1.26	0.63
5	- 8.33	- 0.09	0.31
6	- 0.59	1.66	0.66
7	- 0.47	2.34	0.64
8	- 0.62	1.88	0.67
9	- 0.50	1.97	0.64
10	0.05	0.68	0.49
11	- 0.63	3.81	0.70
12	- 0.57	1.82	0.66
13	- 0.26	3.11	0.58
14	- 0.52	1.73	0.64
15	- 0.39	2.05	0.61
16	0.67	0.92	0.37

The results of this analysis indicated that the items 2 to 15 did not increase gradually in difficulty. Furthermore, the discrimination coefficient of item 5 showed an extremely low value (see section 3.4.1), compared to the value range stated by De Ayala (2009).

Table 32 displays difficulty, discrimination and guessing parameters and proportion of correctly scored items for version 3 of the subtest 'Series'.

Table 32 Subtest 'Series' version 3

Item number	Difficulty coefficient	Discrimination coefficient	Guessing coefficient	Proportion correct
1	- 1.63	2.04	0.00	0.87
2	- 0.69	29.98	0.06	0.74
3	- 0.62	4.05	0.32	0.78
4	- 1.47	0.96	0.00	0.75
5	0 71	2.29	0.32	0.52
6	- 1.00	3.70	0.00	0.78
7	0.44	1.42	0.04	0.40
8	0.59	1.75	0.11	0.40
9	- 0.04	36.38	0.22	0.65
10	- 0.24	4.34	0.00	0.56
11	0.14	2.51	0.19	0.54
12	- 0.05	28.53	0.22	0.65
13	0.04	1.59	0.00	0.47
14	0.15	4.28	0.35	0.63
15	0.00	45.45	0.28	0.62
16	0.66	1.42	0.04	0.35

The results of this analysis indicated that the items 2 and items 5 to 15 did not gradually increase in difficulty. In addition, the discrimination parameter of items 2, 9, 12 and 15 had extreme values and the guessing coefficient of items 3, 5, 14 and 15 showed that a higher than expected number of people guessed the item incorrectly (see section 3.4.1).

Table 33 displays difficulty and discrimination parameters and proportion of correctly scored items for version 1 of the subtest 'Odd One Out'.

Table 33 Subtest 'Odd One Out' version 1

Item number	Difficulty coefficient	Discrimination coefficient	Proportion correct
1	- 1.45	2.12	0.86
2	- 1.79	3.08	0.94
3	- 1.34	24.34	0.88
4	- 1.84	2.24	0.92
5	- 1.35	103.81	0.85
6	- 1.06	4.74	0.82
7	- 1.34	4.09	0.88
8	- 0.86	2.97	0.76
9	- 0.81	2.06	0.73
10	- 0.60	4.44	0.70
11	-.0.24	1.40	0.56
12	0.02	2.73	0.50
13	0.34	1.79	0.41
14	2.99	0.67	0.14
15	12.42	0.25	0.05

The difficulty coefficients and proportion of correctly scored items indicated that the items 1 to 4, and item 7 did not increase gradually in difficulty. In addition, items 3 and 5 showed an extremely high value (see section 3.4.1) for the discrimination coefficient when comparing the values to De Ayala's range for 'good values' between 0.8 to 2.5 (2009).

Table 34 displays difficulty and discrimination parameters and proportion of correctly scored items for version 2 of the subtest 'Odd One Out'.

Table 34 Subtest 'Odd One Out' version 2

Item number	Difficulty coefficient	Discrimination coefficient	Proportion correct
1	- 0.77	4.20	0.83
2	- 1.33	- 4.23	0.06
3	- 0.78	25.06	0.91
4	- 1.15	5.49	0.92
5	- 0.70	36.05	0.86
6	- 0.96	4.14	0.88
7	- 0.76	3.20	0.81
8	- 0.37	4.17	0.70
9	- 0.27	4.26	0.67
10	- 0.39	2.32	0.69
11	0.00	1.20	0.55
12	1.57	1.13	0.22
13	0.23	4.10	0.48
14	5.70	0.31	0.16
15	1.60	0.82	0.27

The results indicated that the items 1 to 7, 10 and 12 to 15 did not increase gradually in difficulty. In addition, items 3 and 5 had extremely high discrimination coefficients (see section 3.4.1) and exceeding the range of 'good values' (0.8 to 2.5) considerably (De Ayala, 2009).

Table 35 displays difficulty and discrimination parameters and proportion of correctly scored items for version 3 of the subtest 'Odd One Out'.

Table 35 Subtest 'Odd One Out' version 3

Item number	Difficulty coefficient	Discrimination coefficient	Proportion correct
1	- 1.53	11.80	0.97
2	- 1.25	4.87	0.89
3	- 1.33	29.63	0.89
4	- 0.70	35.68	0.80
5	- 1.31	30.70	0.88
6	- 1.04	3.84	0.85
7	- 0.90	2.71	0.80
8	- 0.55	3.65	0.72
9	- 0.84	2.67	0.79
10	- 0.33	1.93	0.63
11	4.86	0.28	0.22
12	0.38	2.42	0.42
13	1.29	0.77	0.31
14	7.32	0.28	0.12
15	- 10.27	- 0.22	0.09

The difficulty coefficients and proportion of correctly scored items indicated that the items 2 to 6 and 8, 9 and 11 did not increase gradually in difficulty. Furthermore, items 1 and 3 to 5 showed an extremely high value for the discrimination coefficient (see section 3.4.1), as they exceeded the range of 'good values' between 0.8 and 2.5 hugely (De Ayala, 2009).

For the subtest 'Jigsaw', only the proportion of correctly scored items were computed (see table 36), as the study did not have sufficient numbers of participants for a latent trait model analysis.

Table 36 Proportion of correctly scored items for 'Jigsaw'

Item number	Proportion correct
1	0.93
2	0.93
3	0.93
4	0.40
5	0.33
6	0.33
7	0.33
8	0.33
9	0.20
10	0.27
11	0.13

The results indicated that items 1 to 3 did not increase in difficulty as the proportion of correctly scored items did not decrease. In addition, for version 1, results showed that the proportion of correctly scored items did not decrease for items 5 to 8 and 10 and 11.

The proportions of correctly scored items suggested that although there was an overall gradual increase of difficulty, some items stagnated in their degree of difficulty and did not show sufficient grading. Furthermore, the clear drop of proportion correct between items 3 and 4 indicated a considerable increase in difficulty between these two items.

Interim conclusion

The results of the item analysis identified a number of items with no gradual increase in difficulty. However, all items were within the expected range of difficulty, and no extreme values were detected in the proportion of correctly scored items or the difficulty parameter.

The combination of latent trait model analysis and proportion of correctly scored items provided an indication of the degree of difficulty for each items for the subtests 'Series' and 'Odd One Out'. Based on these results the subtests should be revised. The alterations should be based on the results for each item, but should also take the difficulty of items at equivalent levels in the parallel versions into account. Therefore, it would be necessary to either exchange items to fit them according to their level of difficulty, or to re-design the items (or options in the items for subtest 'Series). As the subtests 'Series' and 'Odd One Out' included more items than the subtest 'Jigsaw', a few, inadequate items could be deleted. The discrimination coefficient should be used as further indicator for mis-fitting items. In addition, item difficulty across versions should be level for each subtest. For the subtest 'Jigsaw', the degree of difficulty was identified using the proportion of correctly scored items. These indicated that although the overall difficulty in each subtest increased gradually, the difficulty of several items stagnated and did not show a gradual incremental rise in difficulty. In addition, the clear drop of proportion correctly scored items from item 3 to item 4 suggests that item 4 is significantly more difficult than item 3. A new item with an intermediate level of difficulty should be developed and added between items number 3 and 4.

As the initial theory of difficulty did not in all items account for the number of required rotations per items, these should be controlled for in the revised version of the subtest.

The results of the item analyses showed that a theory of difficulty is an essential foundation for the development of a test, but cannot entirely account

for the difficulty of each item. There are several reasons, why the empirical difficulty of an item might deviate from the theory of difficulty: Firstly, the theory of difficulty is unlikely to captivate every aspect of increasing difficulty, as, for instance, became evident in the subtest 'Jigsaw', in which the number of rotated pieces was not accounted for in all items in its initial version. Secondly, familiarity with shapes and geometric figures is likely to influence the degree of visual complexity for a participant (Forsythe, Mulhern and Sawey, 2008). Therefore, a more complex looking shape might not correspond with a higher degree of difficulty if it is a familiar object to the participant. Thus, the empirical evaluation of item difficulty is central to the development of a test instrument.

One major limitation of this study was the limited sample size. For a latent trait model analysis, Stone (1992) suggested a sample size of 500 for 20 items and Harwell and Janosky (1991) advised a sample of at least 250 for 15 items to obtain the exact misfit of particular items. Consequently, the latent trait model analysis for the subtests 'Series' and 'Odd One Out' was used in combination with the proportion of correctly scored items, and while items should be interpreted on the basis of the results of both analyses, the proportion of correctly scored items should give more weight to the decisions for alterations than the results of the latent trait model analysis. As a result of the small sample size for the subtest 'Jigsaw' the evaluation of items was based only on proportion of correctly scored items.

The revision of the CCIID should be based on these results to ensure that the levels of difficulty gradually increase, while items at equivalent levels in the parallel versions have the same difficulty level. This revision is described in the section 'Instrument' in the following section.

3.4.2. Reliability of the CCIID

Introduction

As discussed in chapter 2.1, reliability is the degree to which a test achieves repeatability of values or scores (Bartram, 1990) and has to be expressed in a set of reliability studies, which together will provide an estimation of the reliability of an instrument (Kline, 2005). Reliability for the CCIID was evaluated for individuals with ID. The evaluation of reliability in this study included internal consistency, test-retest reliability and inter-rater reliability.

Internal consistency is an indication of the homogeneity of a test. For the assessment of internal consistency, the split-half method is frequently used. The test is split into two halves in order to compare the scores on the items of the first half to the scores on the items of the second half (Kline, 1993). A high correlation confirms that both test halves measure identical concepts. The inter-item correlation is another method of assessing internal consistency and is usually computed with the Cronbach's α (Nunnally & Bernstein, 1994). As the CCIID does not consist of equally difficult items, but of items that increase in difficulty, a split-half method or inter-item correlation is not applicable in this case. Internal consistency for the CCIID was assessed based on the correlations between the parallel versions of the subtests 'Series' and 'Odd One Out'. Internal consistency was not examined for the subtest 'Jigsaw', as it consisted of only one version. Internal consistency was examined using the data obtained for the latent trait model analysis as the analysis required non-adaptively scored test results. Non-adaptive scoring was necessary, as in the adaptive scoring system (described in section 3.2.1), the scores of the parallel versions 2 and 3 would depend on the scores of the previous versions, which, therefore, would have violated the assumption of independence of observations.

As the parallel versions were developed based on the same theoretical foundation, and item revisions were made to construct equivalent item difficulty across the parallel versions, the alternative hypothesis was that *the internal consistency of the CCIID can be confirmed for individuals with ID.*

Test-retest reliability concerns the stability of a test over time and was discussed in chapter 2.1. To ensure that test-retest reliability would not be overestimated due to a lack of internal consistency (see chapter 2) this study investigated test-retest reliability and internal consistency of the CCIID for individuals with ID.

The alternative hypothesis was that *test-retest reliability can be demonstrated for the administration of the CCIID for individuals with ID.*

Inter-rater reliability refers to the influence of the person who administers the test on the test result (see chapter 2). Although the CCIID provides an objective, computerized scoring system and testing is standardized, the test assistant could influence the answer through nonverbal clues or affect the performance in other ways. Consequently, the influence of the test assistant on the test outcome should be assessed. This study examined inter-rater reliability of the CCIID for a population of individuals with ID.

The alternative hypothesis was that *inter-rater reliability can be confirmed for the use of the CCIID for individuals with ID.*

Method

Participants

Participants for the study of internal consistency of the subtests 'Series' and 'Odd One Out' were those included in the initial item analysis of the CCIID (see section 3.4.1.).

The study of test-retest reliability included 27 participants (20 male, 7 female) between 11 and 16 years of age. The participants attended one of two special schools for individuals with moderate and severe learning disabilities.

Participants for the study of inter-rater reliability were 25 table tennis players (13 male, 12 female) between 14 and 46 years of age who had taken part in two different international sports competitions for individuals with ID (INAS European and Open Table Tennis Championships 2008 Ontinyent/Spain and INAS Global Games, 2009 Liberec/Czech Republic).

The studies included participants with different causes for intellectual disability as specific causes of the disability can only be established in a minority of cases (Kaski, 2009).. For all studies, ethical approval had been obtained from the Loughborough Ethics Committee (see appendix B). All participants -and in case of power of attorney, their parents or carers- had been given information about the purpose of the study and gave informed consent prior to the study (see appendix A).

Instrument

Based on the initial assessment of the CCIID, the following items were changed:

Subtest 'Series'
The examples remained the same as in the initial version of this subtest. The following changes were made to the parallel versions:

Version 1:
- For items 3, 4 and 7 the design was changed to make it easier to discriminate between the correct and incorrect options:

- One option was changed, so it would obviously not belong with the other shapes in this 'Series' item.
- One option was changed into a second, obviously wrong option, but one that could belong to this 'Series' item.
- The third and fourth option were more similar, but also changed slightly, so it would be easier to discriminate between the correct and the incorrect options.

- Items 8, 12 and 14 were deleted.
- The designs of items 9 and 10 were changed into shapes similar to the equivalent items in the initial version of version 3, as these showed the appropriate difficulty level.
- The options of item 13 were re-designed to increase the level of difficulty to discriminate between the correct and the incorrect option.

Version 2:

- The designs of items 2 and 6 were changed to a shape similar to those of version 1, as these showed the appropriate difficulty level.
- For items 3 and 4 the options were re-designed to decrease the difficulty to discriminate between the correct and the incorrect option.
- Options for item 5 were changed to increase the difficulty to discriminate between the correct and the incorrect option.
- The design of items 8, 9 and 10 was changed to resemble more closely the concepts of the items used in version 3.
- Item 11 was changed into a shape more similar to the equivalent item in version 2.
- Items 14 to 16 were deleted.

Version 3:

- For items 1, 3, 4, 5 and 6 the options were re-designed to increase the distinction between the correct and the incorrect options.
- Items 8, 12 and 15 were deleted.

- For item 14 the number of distractors was increased in order to increase the level of difficulty.

Subtest 'Odd One Out'
The following revisions were made to the examples and parallel versions:

Examples:
- For examples 2 the correct option ('odd one out') and one shape were changed in order to decrease the difficulty to discriminate between correct and incorrect options.
- For examples 3 the incorrect options were re-designed by increasing their discriminating feature in order to lower the difficulty to discriminate between correct and incorrect options.

Version 1:
- The distinctiveness of the correct option was increased in items 3, 4, 5 and 7 to decrease the level of difficulty for their options.
- Item 11 was deleted.

Version 2:
- For items 2 and 5, the correct option was re-designed to be more distinctive from the incorrect options in order to make it easier to discriminate between correct and incorrect options.
- The sequence of several items was changed according to their difficulty coefficient (see chapter 3.4.1): Items 3 and 4 were switched, items 8, 9 and 10 were switched and items 14 and 15 exchanged.
- Item 12 was deleted.

Version 3:

- Items 2 and 4 were re-designed in order to increase the distinctiveness of the correct option.
- Items 3 and 5 and items 8 and 9 exchanged positions according to their difficulty coefficients (see section 3.4.1).
- Item 11 was deleted.

Subtest Jigsaw:

The examples remained the same, but the following changes were made to the test items:

- Item 2 was deleted as all people who scored item 1 correctly also scored item 2 correctly, therefore it was superfluous.

- A new item was added between items 3 and 4, which contained one patterned piece that did not need to be rotated to complete the item. This was done to create an intermediate level of difficulty between items 3 and 4, as many lower ability participants had no problems with item 3 but could not use turn-able pieces as needed for item 4.

- Numbers of pieces that needed rotation to complete the item was not controlled for all items in the initial version of the subtest. This was corrected in the revision as follows:

    - Items 1 to 3: not corrections as there were no turned pieces used
    - Item 4: only 1 piece to rotate
    - Item 5: 2 pieces to rotate
    - Item 6: 3 pieces to rotate
    - Item 7: 4 pieces to rotate
    - Item 8: 5 pieces to rotate
    - Item 9: 6 pieces to rotate
    - Item 10: 7 pieces to rotate
    - Item 11: 8 pieces to rotate

All revised items can be found in appendix E. Internal consistency was investigated using the non-adaptive scoring form of the CCIID (all items of each version were administered regardless of how many the participant scored incorrectly), test-retest and inter-rater reliability of the CCIID were evaluated using the adaptive scoring system described in section 3.2.1.

Test procedure and environment

For internal consistency, of the subtests 'Series' and 'Odd One Out' test procedure and environment were discussed in section 3.4.1.

For test-retest reliability, participants were assessed in their respective schools in an unoccupied, quiet office space. The test battery was administered by the same test assistant on both occasions (author of this thesis). Participants of both schools were retested after a period of three months by the same test assistant. A three month interval was chosen to minimize the effects of memory (Ghiselli, Campbell & Zedek, 1981) and learning effects (Kline, 2005).

Inter-rater reliability was assessed during sport competitions with different test assistants (author of this thesis and trained final year psychology student or clinical psychology students/researchers) testing the same participants on two occasions with an interval of 8 months. On both occasions, assessment took place in a quiet, empty room with sufficient space in order to test two participants simultaneously without disturbing each other's performance.

Testing took between 15 and 25 minutes per participant depending on ability. The subtests were administered in the sequence: 'Series', 'Odd One Out' and 'Jigsaw'.

Statistical analysis

Internal consistency of CCIID was evaluated based on the scores of the parallel versions of the subtests 'Series' and 'Odd One Out' using correlations.

Test-retest and inter-rater reliability of the CCIID were estimated using scores on the three subtests separately as well as the overall composite score, which was based on the z-scores of the three subtests from all participants included in the reliability and validity studies (sum of z-scores/3). Test-retest and inter-rater reliability were assessed comparing the scores obtained in the first and second testing session using Spearman's rank correlations. A non-parametric Wilcoxon signed rank-test was used to investigate possible learning effects between the first and second test administration. Prior to all analyses, descriptive statistics were computed for all analysis and assumptions were tested. The level of significance was 0.05. All data were analysed using SPSS version 14.0.

Results

For the estimation of internal consistency, correlations between the parallel versions of the subtests 'Series' and 'Odd One Out' were computed. An examination of the distribution of scores did not confirm the assumptions of normality. Therefore, it was decided to assess internal consistency using Spearman's rank correlations. Descriptive statistics for the study of internal consistency are displayed in table 28, chapter 3.4.1. Correlations between the parallel versions are displayed in table 37.

Table 37 Correlations between parallel versions of each subtest (Spearman's rho)

Subtest	Correlations between series	Rho	N
'Series'	A and B	0.84**	65
	A and C	0.81**	65
	B and C	0.75**	66
'Odd One Out'	A and B	0.77**	62
	A and C	0.73**	63
	C and B	0.83**	60

**Correlation is significant at the 0.01 level

The results showed large correlations between the parallel versions of the subtests 'Series' and 'Odd One Out'. Therefore, these results supported the alternative hypothesis that internal consistency of the CCIID can be confirmed for individuals with ID.

Test-retest reliability was evaluated using the scores of all subtests separately, as well as the CCIID composite score (= CCIID score). Descriptive statistics are displayed in table 38.

Table 38 Descriptive statistics

	N	Min. score	Max. score	Mean	SD
Age	27	11	16	13.56	1.55
'Series' 1st assessment	27	0	38	13.44	11.40
'Series' 2nd assessment	27	0	32	14.07	9.89
'Odd One Out' 1st assessment	26	0	37	23.85	9.95
'Odd One Out' 2nd assessment	26	4	34	24.15	8.43
'Jigsaw' 1st assessment	26	0	10	3.00	2.98
'Jigsaw' 2nd assessment	26	0	11	3.65	3.46
CCIID score 1st assessment	25	- 1.61	1.86	- 0.05	1.05
CCIID score 2nd assessment	25	- 1.53	1.50	0.02	0.84

Again, an inspection of the distributions did not support the assumption of normality. Consequently, the correlations were computed using Spearman's rho. Correlations are displayed in table 39.

Table 39 Correlations between test and retest scores for each subtest and the CCIID score using Spearman's rho

	Rho	N
'Series'	0.84**	27
'Odd One Out'	0.77**	25
'Jigsaw'	0.82**	26
CCIID score	0.88**	24

Note: **is significant at the 0.01 level

The results showed large correlations between the test and retest assessment for the CCIID. Therefore, the results supported the alternative hypothesis that test-retest reliability can be demonstrated for the administration of the CCIID for individuals with ID.

Furthermore, the results of a Wilcoxon signed rank test for each subtest, as well as the CCIID composite score showed that there were no significant differences between the scores of the first and second assessment, which indicated that there was no significant learning effect between the two test administrations.

Inter-rater reliability was investigated using the scores of all subtests separately as well as a CCIID composite score. Descriptive statistics are displayed in table 40.

Table 40 Descriptive statistics

	N	Min. score	Max. score	Mean	SD
age	25	14	46	26.80	8.63
'Series' 1st assessment	24	7	35	16.64	7.02
'Series' 2nd assessment	24	3	33	16.92	7.29
'Odd One Out' 1st assessment	24	10	36	27.09	5.39
'Odd One Out' 2nd assessment	24	4	38	29.52	6.78
'Jigsaw' 1st assessment	25	2	11	5.20	2.59
'Jigsaw' 2nd assessment	25	2	11	5.72	2.73
CCIID score 1st assessment	22	- 0.27	1.51	0.43	0.54
CCIID score 2nd assessment	22	- 2.35	1.38	0.00	0.84

An inspection of the distribution of scores (subtests and CCIID score) did not support the assumption of normality. Consequently, the correlations were computed using Spearman's rho. Correlations are displayed in table 41.

Table 41 Correlations between scores obtained from two different test assistants using Spearman's rho

	Rho	N
'Series'	0.58**	24
'Odd One Out'	0.42 *	24
'Jigsaw'	0.76**	25
Composite score CCIID	0.83**	22

Note: *is significant at the 0.05 level (2-tailed) ; **is significant at the 0.01 level (2-tailed)

The analysis revealed significant correlations between the test administrations of two different test assistants for all subtests with a large correlation for the CCIID composite score. This supported the alternative hypothesis that inter-rater reliability can be confirmed for the use of the CCIID for individuals with ID.

In order to examine learning effects, a Wilcoxon signed rank test was computed for each subtest and the CCIID composite score. The results demonstrated that there were no significant differences between the scores of the first and second assessment, which indicated that there were no significant learning effects between the two test administrations.

Interim conclusion

The study supported the overall reliability of the CCIID. The investigation of internal consistency for the subtests 'Series' and 'Odd One Out' revealed that the parallel versions were highly correlated. The data reflect estimates of

internal consistency seen in other intelligence tests, such as the Wechlser Intelligence Scale for Children - IV (Wechsler, 2003) and nonverbal tests such as the CTONI (Cohen & Spenciner, 2003), the Leiter-R (Roid & Miller, 1997; 1999) and the SON-R 5½ -17 intelligence test (Tellegen & Laros, 1993). The evaluation of test-retest reliability showed large correlations between the scores of the first and second test administration for all three subtests, as well as for the CCIID composite score, and is also similar to other commonly used intelligence tests, such as the Stanford-Binet Intelligence Scale for the use for individuals with ID (retest reliability for abstract/visual reasoning; r = .90; Dacey, Nelson & Stoeckel, 1999).

Furthermore, the evaluation of inter-rater reliability showed significant correlations between scores obtained from different test assistants. Correlations between the first and second assessment for the subtests 'Series' and 'Odd One out' were not as large as for test-retest reliability, which might be a result of one of several factors. Firstly, compared to the subtest 'Jigsaw', the chance to score an item correctly is higher for these two subtests as they are multiple choice tests. This might influence the stability of the scores. However, using a scoring system based on the three parallel versions, the influence of chance on the results should be minimal. Secondly, although the test assistants might have felt confident that the participant understood the instructions, this might not have been the case. Further research should investigate to what extent this confidence in the comprehension of the instructions is justified. Thirdly, some participants had competitions on the day of the second assessment. Sports competitions are physically and psychologically highly demanding situations, which are likely to affect the cognitive performance of the athlete during the recovery period. This might have influenced their inductive reasoning abilities more than their visual processing abilities. Therefore, a longer recovery period for the athletes should be included in future assessments and athletes should not be tested on the day of their competitions. However, the large correlation between the

composite CCIID scores of the first and the second assessment clearly supports the inter-rater reliability of the test battery.

The study had several limitations. Firstly, internal consistency was assessed using the initial version of the CCIID, which was scored in a non-adaptive form (i.e. all items of each version were administered regardless of how many the participant scored incorrectly). Due to practical restrictions, non-adaptively scored data could not be obtained again for the revised CCIID. However, the revision was based on proportions of correctly scored items, and the results of the latent trait model analysis and included only a limited number of items. Therefore, little change in internal consistency can be expected for the revised version, and, as several items were changed to be more similar across parallel versions of the subtests, any change in internal consistency would be likely to be an improvement to the current results. Furthermore, internal consistency could not be investigated for the subtest 'Jigsaw' as it did not consist of parallel versions. However, test-retest reliability is unlikely to be affected by that, as 'Jigsaw' is not a multiple choice test and, therefore, it is doubtful that it would be scored correctly by chance (see section 3.2.4.).

Secondly, occasional interruptions did occur during the administration of the test battery, but participants did not seem unduly affected in their attention towards the assessment.

Thirdly, due to the relatively small sample size, reliability could not be investigated in different subgroups (for age, level of intellectual functioning or origin of disability). However, reliability was confirmed for the use of the CCIID for the diverse population of this sample, which included individuals functioning on different levels, ID of different origins and an age range from 11 to 44 years of age. In addition, the reliability of similar subtests for inductive reasoning and visual processing abilities does not seem to vary greatly in different clinical groups or for different levels of intellectual disability (Wechsler, 2003).

171

Based on these results it can be concluded that the CCIID is a reliable instrument for individuals with ID to assess *inductive reasoning* and *visual processing* abilities. Reliability is necessary, but not sufficient for validity (Kline, 1993). Therefore, it is very important to establish both reliability and validity, for any psychometric test. The following chapter will examine validity of the CCIID for the individuals with ID.

3.4.3. Validity of the CCIID

The validity of an instrument determines to what extent the instrument measures what it is supposed to measure (American Educational Research Association, 1999). As discussed in chapter 2, validity for psychological tests cannot be expressed in a single value, but needs the accumulation of evidence supporting validity (Aiken, 1994; Nunnally & Bernstein, 1994). There are several methods of validity that will be investigated to evaluate the overall validity of the CCIID:

Content validity is a theoretical consideration that refers to the extent to which the items of an instrument represent the concept which the instrument measures. The content validity of the CCIID has been researched at the beginning of its development (see chapter 2) and the theoretical foundation supported the validity of the CCIID for the use for individuals with ID from different cultural backgrounds.

Construct validity refers to the relationships between scores within a test (see chapter 2). The construct validity of the CCIID was assessed with a principal component analysis. Similar nonverbal test batteries, such as the SON-R, showed one dominant factor (see section 2.2.2) when assessing construct validity for individuals with ID. Therefore, the alternative hypothesis was that

construct validity will be confirmed for the use of CCIID for the assessment of individuals with ID.

<u>Criterion validity</u> refers to the extent to which a test predicts or correlates to a certain criterion (see chapter 2). Criterion validity of the CCIID will be evaluated in a comparison of two separately obtained sets of IQ scores. The first set consists of scores on the Wechsler Adult Intelligence Scale (WAIS), which were established outside of the context of this study. The participants had obtained these scores to prove eligibility for international sports events in the category 'Intellectual Disability'. The second set consisted of IQ scores obtained through the assessment with the SON-R intelligence test for studies 1 to 3 (see chapter 2 of this thesis). As the CCIID is based on subtests used in both intelligence tests, the SON-R and the Wechsler Scales, the second alternative hypothesis was that *there is an association between the scores of the CCIID and the scores on Wechsler Adult Intelligence Scale as well as scores on the SON-R for individuals with ID.*

The CCIID assessed, similar to the SON-R, inductive reasoning and visualisation processing abilities. Both test batteries excluded subtests for verbal abilities or general knowledge which in contrast, are part of the Wechsler Adult Intelligence Scale (Wechsler, 1997). Consequently, the third alternative hypothesis was that *the correlation between scores on the CCIID and scores on the SON-R is significantly higher than between scores on the CCIID and scores on the Wechsler Intelligence Test*

Method

Participants

The evaluation of <u>construct validity</u> included 91 participants with ID (60 male, 35 female) between 6 and 52 years of age. For the background of the participants see table 42.

Table 42 Background of participants for construct validity CCIID

N	Background participant
42	2008 INAS European Table Tennis Championships for individuals with ID Ontinyent/Spain
8	Aspirants for national training squad invited by MENCAP
7	Members of a special needs afterschool and youth club
34	Pupils of special needs schools for different levels of intellectual disability
4	Members of the national table tennis training squad in the category 'Intellectual Disability'

The assessment of <u>criterion validity</u> included 37 participants in total. Eighteen table tennis players (14 male, 4 female) with an age range between 18 to 52 years had scores on the CCIID and the Wechsler Adult Intelligence Scale. All 18 were participants in the 2009 INAS - European Table Tennis Championships in Ontinyent/Spain. In addition, 11 table tennis players who were training for the English table tennis team in the category 'Intellectual Disability' and 8 aspirants for the national training squad (11 male, 8 female) between 14 and 45 years of age had scores on the CCIID and IQ scores on the SON-R 5½ -17 intelligence test.

For all studies, ethical approval had been obtained from the Loughborough Ethics Committee (see appendix B). Athletes and -in case of power of attorney- their parents or carers had been given information about the purpose of the study and gave informed consent prior to the assessment (see appendix A). The studies included participants with different causes for intellectual disability as specific causes of the disability can only be established in a minority of cases (Kaski, 2009).

Instrument

Construct and criterion validity were investigated using the adaptive test procedure (see section 3.2.1.) of the revised version of the CCIID.

Test environment and procedures

All tests were administered by trained test assistants (author of this thesis and a trained final year/research psychology student), and participants were given time to familiarize themselves with the test environment. The assessment with CCIID took place in large, quiet rooms. On two occasions (2008 INAS European table tennis championships and national training day for MENCAP) two test assistants administered the test battery in the same room. In these cases, the rooms were chosen with ample space between testing stations to ensure that participants could be tested simultaneously without disturbing each other's performance. Testing took between 15 and 25 minutes per participant depending on ability. The subtests were administered in the sequence: 'Series', 'Odd One Out' and 'Jigsaw'. All athletes (table tennis players and aspirants for training squad) were tested during sports events. On each occasion, the testing stations were in separate rooms from the main event. In the schools, the tests were administered in empty offices. At the special needs afterschool and youth club the CCIID was administered in a part of the common room that was screened off.

Statistical analysis

Construct validity was assessed using the scores on the subtests 'Series', 'Odd One Out' and 'Jigsaw' separately. Additionally, the composite CCIID score based on the sum of the z-scores of these three subtests (see section 3.4.2), and existing IQ scores (Wechsler Adult Intelligence Scale and scores on the SON-R) were used to assess criterion validity. Construct validity of CCIID for the assessment of individuals with ID was examined using an exploratory principal component analysis with the scores of the three subtests 'Series', 'Odd One Out' and 'Jigsaw'. All 91 participants had scores on the subtest 'Series', 88 participants had scores on the subtest 'Odd One Out' and 86 participants had scores on the subtest 'Jigsaw'.

In order to establish criterion validity for the SON-R for use of the CCIID for individuals with ID, Spearman's rank correlation was calculated between the composite score of the CCIID and the registration IQ score of 18 participants. Additionally, Spearman's rank correlations were computed between the composite score of the CCIID and the IQ scores on the SON-R of 19 participants. Subsequently, the statistical significance of the difference between the correlation coefficients was tested using Fischer's z-transformation.

Prior to all analyses, descriptive statistics were computed for all analysis and assumptions were tested prior to the exploratory principal component analysis as described in section 2.1.4.

Data were analysed using SPSS (version 14.0) and the level of significance was 0.05.

Results

Construct validity

The average age of all 91 participants was 21.88 (SD 10.68) years, with a range from 6 to 52 years of age. None of the participants reached ceiling scores on the subtests 'Series' or 'Odd One Out'. Two participants reached ceiling scores on the subtest 'Jigsaw'. One participant performed at floor in the subtest 'Series', two participants obtained zero scores in the subtest 'Odd One Out' and nine participants performed at floor in the subtest 'Jigsaw'. The means, standard deviations and minimum and maximum scores of the subtests for all participants are displayed in table 43.

Table 43 Descriptive analysis of subtests for construct validity

Subtest	Minimum	Maximum	Mean	SD	N
'Series'	0	38	12.95	8.47	95
'Odd One Out'	0	37	24.46	8.43	92
'Jigsaw'	0	11	3.46	2.81	90

In order to study the construct validity of the CCIID for individuals with ID, the scores on the three subtests were entered into a principal component analysis after the suitability of the data was assessed see section 2.1.4.). The scores on the subtests showed a normal distribution. The correlation matrix revealed that all coefficients were .30 and above. The Kaiser-Meyer-Oklin value was .62 and, therefore, above the recommended value of .6 (Kaiser, 1974). Bartlett's Test of Sphericity reached statistical significance. These indicators suggested that the data was suitable for factor analysis (Field, 2005). The principal component analysis established one component with an Eigenvalue of 2.04, explaining 67 % of the variance (see table 44).

Table 44 Results principal component analysis

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cum. %
1	2.04	67.85	67.85	2.04	67.85	67.85
2	0.67	22.34	90.19			
3	0.30	9.82	100.00			

A parallel analysis confirmed the dominance of one factor: Only one component exceeded the Eigenvalue of the corresponding criterion value in a randomly generated data matrix of the same size.

The dominant factor is regarded to be the underlying intelligence factor g (Carroll, 1993). Furthermore, the results revealed that all subtests loaded highly in this dominant factor (see table 45).

Table 45 Component matrix

	Component 1
subtest 'Series'	.852
subtest 'Odd One Out'	.712
subtest 'Jigsaw'	.896

These findings supported the alternative hypothesis that construct validity for the CCIID can be confirmed.

Criterion validity

Participants included in the study of criterion validity had an average age of 33.00 years (SD 9.54), with a range from 18 to 52 years of age. The means,

standard deviations and minimum and maximum scores of the CCIID composite score, scores on the Wechsler Intelligence Scale for Adults-III (WAIS-III) and SON-R scores are displayed in table 46.

Table 46 Descriptive analysis of scores used for criterion validity

Test	Minimum	Maximum	Mean	SD	N
CCIID scores	-1.81	1.39	-.08[1]	0.85	37
WAIS-III scores	40	74	60.44	9.72	18
SON-R scores	48	72	53.60	6.64	19

Results showed a significant correlation between the CCIID scores and scores on the WAIS-III (r_s= .66, p< .01, n= 18) as well as between CCIID scores and scores on the SON-R (r_s= .82, p< .01, n= 19). Therefore, the results confirmed the second alternative hypothesis that there is a significant association between the scores of the CCIID and the scores on Wechsler Adult Intelligence Scale as well as scores on the SON-R for individuals with ID.

Subsequently, the two correlations were statistically compared based on the observed z-value (Howell, 2002). The outcome revealed that there was no significant difference between the association of the CCIID scores with the WAIS-III scores and the association between CCIID scores and the SON-R scores. The third alternative hypothesis, that the correlation between scores on the CCIID and scores on the SON-R is significantly higher than between scores on the CCIID and scores on the Wechsler Intelligence Test, was thus not accepted.

---

[1] Negative value, as only a part of sample that was used to compute CCIID scores, was included to establish criterion validity

Interim conclusion

Both construct and criterion validity supported the overall validity of the CCIID to assess cognitive functioning in individuals with ID.

The results of the investigation of construct validity confirmed one underlying component, which is regarded to be a general intelligence factor "g" (Carroll, 1993; Gustafsson, 1988). Although the CCIID measures two separate cognitive abilities (fluid reasoning and visual processing abilities), it can be expected that performance on both abilities showed one underlying factor. Marshalek, Lohman and Snow (1983) showed that both abilities are often difficult to differentiate. This is reflected in the radex model developed by Snow, Kyllonen and Marshalek (1984), which supports the centrality of inductive reasoning tasks and visual processing abilities for the general intelligence factor "g". This model maps inter-correlations between tests as distances, with tests for inductive reasoning and visual processing being close to the centre which represents "g".

The validity of the CCIID for individuals with ID is further confirmed by the examination of criterion validity. As expected, the outcome showed a significant relationship with both intelligence tests (WAIS and SON-R). However, although the association between scores on the CCIID and the SON-R was stronger than between scores on the CCIID and the WAIS, this was not a statistically significant difference. Nevertheless, results indicate that the CCIID measures a similar concept to both other intelligence tests.

One limitation of the study was the sample size, particularly for the study of construct validity. The recommended sample size is 300 or more participants in order to ensure a stable factor solution (Field, 2005) for a principal component analysis. However, the Kaiser-Meyer-Olin measure of sampling adequacy in this study (KMO = 0.62) suggested that the sample size was acceptable for this principal component analysis (Kaiser, 1974).

A larger sample size in the study of criterion validity might also have revealed a significant difference in the correlations between the scores on the CCIID and scores on the WAIS and scores on the CCIID and scores on the SON-R. Furthermore, similar to our earlier work (see chapter 2.3), occasional interruptions during the assessments occurred, but this did not seem to unduly affect the attention of the participants.

In sum, these studies clearly supported the validity of the CCIID for individuals with ID. The outcome of the validity studies confirmed that the CCIID measures what it set out to measure and that it is an appropriate assessment tool for intellectual functioning in individuals with ID.

3.4.4. Discussion

The results confirmed the validity and reliability of the CCIID for the assessment of cognitive abilities in individuals with ID. Psychometric properties were assessed based on a sample with a wide range of cognitive functioning. Testing at schools for individuals with moderate and severe ID showed that the CCIID is also suitable for the assessment of individuals with a high degree of intellectual disability. None of the participants of the overall sample reached ceiling scores in the subtests 'Odd One Out' and 'Series' and only two participants reached ceiling in the subtest 'Jigsaw'. None of the participants had zero scores in all three subtests. Four participants had zero scores in either the subtests 'Odd One Out' (2 participants) or 'Series' (2 participants). Nine participants obtained zero scores for the subtest 'Jigsaw' and two participants could not complete the subtest 'Jigsaw' as their physical coordination ability did not allow them to move the pieces into the correct position when supported during the examples. This indicates that, although the test battery is suitable for a wide range of degree of intellectual disability, physical restrictions (which might be associated with the intellectual disability)

181

might limit the application of the test battery and this should be taken into account when administering the tests.

The *multi mode approach to the instructions and use of the practice* items proved to be a very satisfactory means to control the understanding of the task before commencing the assessment. This control is particularly important to ensure that the test battery tests the cognitive abilities it was designed to assess, and not the comprehension of the instructions. The *approach to the instructions was suitable for English speaking participants*, as well as to the international participants tested during the 2008 INAS European and Open Table Tennis Championships in Ontinyent/Spain. In all cases the test assistants felt confident that the participant had understood fully the instructions before the assessment started.

The discrepancy between the British and DSM-IV standards for intellectual disability is expected to be minimal as both standards use the same international criteria except the determination of limitations of intellectual impairment can be based for British standards on the judgement of psychologists instead of an IQ score as it is defined by the DSM-IV. Possibly, the a British sample might be biased towards individuals with a higher degree of intellectual disability, as they attend special schools for learning disabilities, while children and teenagers with mild learning disabilities are often not diagnosed and attend mainstream schools. Another effect of the discrepancy might be the inclusion of other disorders, which lead to learning problems, such as attention deficit hyperactivity disorder or autistic spectrum disorders. Overall, however, the CCIID provided enough discrimination towards both ends of the scale to assess the cognitive functioning of all participants included in the validity and reliability studies.

Further research should establish the degree of cultural fairness of the CCIID. The test battery had been developed based on the criteria for test fairness for nonverbal intelligence tests (see section 3.1.4.) and, so far, all participants

182

seemed to understand the function of the shapes, and no comments were made by participants or coaches that indicated a cultural bias in the comprehension of the subtests or items. However, also nonverbal intelligence tests are not entirely free from cultural bias (Sattler, 1992; Rosselli & Ardila, 2003). Therefore, the degree of cultural bias should be further examined.

In sum, the CCIID proved to be a short, user-friendly instrument which is very suitable for the target population. Further research with the CCIID should investigate the association between cognitive abilities and sports performance for athletes with ID. As floor and ceiling effects of the test battery were limited, the CCIID showed excellent discrimination properties at the lower end of cognitive abilities. Therefore, the examination of the relationship between cognitive abilities and sports performance with the CCIID will provide more accurate results than studies using conventional IQ tests. In addition, the CCIID could be used in a wider context when investigating cognitive abilities of individuals with ID. Possible applications for the CCIID are as a tool for talent spotting as visual processing and inductive reasoning appear to be predictive of sports performance in certain sports disciplines for individuals with ID, or as a diagnostic instrument e.g. in schools to identify areas of problematic development, which then could get special attention. Additional research in these areas will have to determine the benefits of the test battery.

Chapter IV: Sport-specific studies

4.1. Introduction

The purpose of the following studies was to confirm the association between cognitive abilities and sports performance for elite athletes with ID, using the Computerized Cognitive test battery for Individuals with Intellectual Disabilities (CCIID).

The initial studies using the nonverbal SON-R intelligence test had revealed a significant association between cognitive abilities and sports competition performance for table tennis players, but not for swimmers. Track and field athletes did not have competition scores in the initial studies. However, as discussed in chapter 2.3, the intelligence test used in these studies had a number of limitations, such as floor effects and cultural bias. Subsequently, the CCIID was developed, in order to obtain a more accurate picture of the relationship between cognitive abilities and sports performance for elite athletes with ID.

The following three studies investigated the relationship between cognitive abilities, as measured with the CCIID, and sports competition performance for the sports disciplines table tennis, track athletics and swimming.

Based on results of the initial studies, the alternative hypothesis for table tennis players is:

There is a positive association between the scores on the CCIID and sports competition performance for elite table tennis players with ID.

There is no equivalent in the CCIID to the subtest 'Categories' of the SON-R, which was a significant predictor for table tennis performance. Therefore, both

the CCIID composite scores, as well as the scores on the three subtests separately, will be investigated in their relationship to sports performance.

Again, as a consequence of the initial studies, the alternative hypothesis for track athletics and swimming is:

There is no association between the scores on the CCIID and sports competition performance for either elite track athletes or swimmers with ID.

The CCIID composite scores as well as the scores on the three subtests separately were used to examine the relationship between cognitive abilities and sports performance.

4.2. Methods

4.2.1. Participants

All participants were elite athletes with ID competing in international sports events. For table tennis, participants took part in the 2008 INAS European and Open table tennis championships in Ontinyent/Spain, while swimmers and track athletes were participants in the Global Games 2009 in Liberec/Czech Republic. Their intellectual disability was established prior to the competition and based on the criteria of the definition of the World Health Organisation: To be eligible for international competitions, the athletes needed to have, firstly, an IQ score of 75 or below; secondly, significant limitations of adaptive behaviour; and thirdly, the disability needed to be evident before the age of 18. Table 47 summarizes the gender distribution and age of the participants for the sport-specific studies.

Table 47 Distribution for gender and age for sport-specific studies

	male	female	age	n
Table tennis	28	14	14 - 42	42
Track athletes	59	25	13 - 39	84
Swimming	57	35	14 - 35	92

For all studies athletes and their parents or carers had been given information about the purpose of the study and gave informed consent prior to the assessment (see appendix A and G). Ethical approval had been obtained from the Loughborough Ethics Committee (see appendix B). The studies included participants with different causes for intellectual disability as specific causes of the disability can only be established in a minority of cases (Kaski, 2009).

4.2.2. Instruments

The degree of cognitive ability was measured using the revised version of the CCIID (see section 3.4.2) with the adaptive scoring form as described in section 3.2.1.

4.2.3. Test environment and procedures

Table tennis

Athletes were assessed by one of two test assistants (author of this thesis and a trained final year psychology student) in a quiet, empty office space adjacent to the Ontinyent Communal sports hall where competitions took place. The room offered ample space between testing stations to test two participants simultaneously without disturbing each other's performance. Coaches were present to support the instructions during the examples if necessary. Due to training and competition schedule, some participants had physical exercise before the assessment but all participants had ample rest prior to the assessment.

Track athletes and swimmers

Assessment took place during the Global Games 2009 in Liberec/Czech Republic in vacant flats in the halls of residence of Liberec University. Two testing stations were located in each room, which offered enough space to ensure that participants would not disturb each other when being tested simultaneously. The tests were administered by trained final year and clinical psychology students, psychology researchers and the author of this thesis. Although some participants had sports competitions earlier on the day of the assessment, they all had ample rest before testing commenced. The CCIID was administered as part of a test battery including a wider range of cognitive abilities. The overall assessment took about 45 minutes.

All participants were given time to familiarize themselves with the environment and the equipment. The purpose of the assessment and the option to stop the assessment at any time was explained again to the participants, if necessary with the help of the coach. Subsequently, the assessment started. For all assessments, the tests were administered in the same test order.

4.2.4. Statistical analysis

Cognitive abilities were measured using the scores on the CCIID subtests separately, and using a composite score, which was based on the z-scores of the three subtests from all participants included in the sport-specific studies (sum of z-scores/3). Reliability of scores was assessed based on comments noted by the test assistants for each participant. These comments included interruptions, clear lack of motivation to complete the tests accurately, a lack of comprehension of test instructions (e.g. on some occasions the coach was not present to translate) and any other occurrence that seemed to influence

the participant's performance. In addition, the scores on the subtests 'Series' and 'Odd One Out' of several participants had to be rated as unreliable due to a fault in the computer program. Subsequently, all unreliable scores were removed prior to the analysis. Therefore, not all participants had scores on all subtests and not for every participant a CCIID score could be calculated.

All sports performance scores were derived from international sports competitions:

- Table tennis performance was calculated using the results of the 2008 INAS European and Open table tennis championships in Ontinyent/Spain. Scores of all sets played by a participant were added up, and the total was divided by the number of sets played by each athlete.

- Sports performance scores for track athletes and swimmers were based on the competition results of the Global Games 2009 in Liberec/Czech Republic. Performance scores for track athletes and swimmers were computed using final times of to construct linear mixed effects models accounting for the speed of each athlete whilst taking into account swimming or running style, distance, distance^2 (to model non-linear effects of distance on speed), age and gender. The model provided a performance score for each athlete based on best times per distance, which was used as the sports performance outcome variable.

Hierarchical linear regression analyses were conducted with sports performance score as dependent variables and using the composite CCIID scores as an independent variable while controlling for sex and age to investigate the association between cognitive abilities and sports performance for elite athletes with ID. Subsequently, hierarchical (stepwise) regression analyses were computed using the scores on the CCIID subtests separately as independent variables, in order to find which subtest predicted physical performance most accurately, while controlling for age and sex. Prior to all

188

analyses, descriptive statistics were calculated and assumptions were tested as described in section 2.1.4. A level of significance of 0.05 was used (two-sided). An a-priori power analysis for the sport-specific study for table tennis competition performance was based on the result of the regression analysis in section 2.2.1., which used a subtest for inductive reasoning as predictor for table tennis competition. The result had shown a medium effect size. Based on that, the a-priori power analysis showed a required sample size of 76 participants (Cohen, 1992). The current study included only 42 participants. The consequences of this will be considered in the discussion of the results. Power levels were calculated retrospectively for swimmers and athletics as no prior studies had established observed effect sizes or variances for competition performance that could be used to calculate power a priori. For the evaluation of results, a power level of 0.80 was regarded as sufficient (Pallant, 2005; Field, 2005). All analyses were conducted in SPSS 16.0.

4.3. Results

4.3.1. Sport-specific study 1: Table tennis

This study included 42 table tennis players (28 male, 14 female). Table 48 shows minimum and maximum scores, mean and SD for the participants of this study.

Table 48 Descriptive statistics table tennis study

	Min.	Max.	Mean	SD	n
age	14	52	29.10	9.46	42
CCIID composite score	- 2.19	1.06	- 0.31	0.77	38
'Series' scores	2	27	13.41	6.63	41
'Odd One Out' scores	1	36	24.83	7.90	40
'Jigsaw' scores	0	11	4.46	2.96	41
table tennis performance scores	2.00	11.08	7.39	2.42	42

All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

A hierarchical regression analysis revealed that CCIID composite scores showed a significant association with table tennis performance when controlling for sex and age (R^2= .23, beta = .30, p<.05). However, CCIID scores were not a significant predictor in this model, and R^2 did not change significantly in comparison to a model only including sex and age. Table 49 shows the model summary of the hierarchical regression analysis for table tennis performance.

Table 49 Summary of Hierarchical Regression Analysis for table tennis performance (N = 38)

Variable	B	SE B	β	p
Step 1				
Age	-.05	.04	-.21	.19
Sex	1.40	.80	.28	.09
Step 2				
Age	-.03	.04	-.12	.49
Sex	1.46	.77	.29	.07
CCIID scores	.93	.49	.30	.07

Note. R^2 = .15 for Step 1; ΔR^2 = .08 for Step 2 (ps= .07).

As discussed in section 2.1.4, the assumptions for hierarchical regression analysis were assessed: tolerance and VIF were both checked and indicated that assumptions for multicollinearity were met. An inspection of the standardized residual scatter plot confirmed normality, linearity,

homoscedasticity and the absence of multivariate outliers (Tabachnick & Fidell, 2007).

When a stepwise hierarchical regression analysis was computed, entering the subtests separately and controlling for sex and age, the results showed that scores on the subtest 'Series' significantly predicted table tennis performance (R^2 = .25, beta = .32, p<.05). Table 50 shows the model summary of the hierarchical regression analysis for table tennis performance.

Table 50 Summary of Hierarchical Regression Analysis for table tennis performance (N = 41)

Variable	B	SE B	β	p
Step 1				
Age	- .05	.04	- .21	.19
Sex	1.39	.77	.28	.09
Step 2				
Age	- .04	.04	- .17	.27
Sex	1.45	.76	.29	.07
scores on the subtest 'Series'	.12	.06	.32	.04*

Note. R^2 = .15 for Step 1; ΔR^2 = .10 for Step 2 (ps<.05). * p< .05

Again, assumptions for hierarchical regression analysis were examined as described in section 2.1.4: tolerance and VIF were both checked and indicated that assumptions for multicollinearity were met. An inspection of the standardized residual scatter plot confirmed normality, linearity, homoscedasticity and the absence of multivariate outliers (Tabachnick & Fidell, 2007).

Therefore, the alternative hypothesis that there is a positive association between the scores on the subtest 'Series' and sports competition performance for elite table tennis players with ID could be confirmed.

4.3.2. Sport-specific study 2: Track athletics

The second pilot study, which investigated track athletes, included 84 participants (59 male, 25 female). Table 51 shows minimum and maximum scores, mean and SD for the participants of this study.

Table 51 Descriptive statistics track athletics study

	minimum	maximum	mean	SD	n
Age	13	39	23.54	4.94	84
CCIID composite score	-1.94	1.75	-.06	.90	50
'Series' scores	3	36	15.00	8.92	52
'Odd One Out' scores	3	41	26.45	8.43	53
'Jigsaw' scores	1	11	4.68	2.79	82
track performance scores	4.73	6.70	5.77	0.44	84

All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

The results of a hierarchical regression revealed that there was no significant association between the track performance scores and the CCIID scores, or the subtest scores, when controlling for sex and age. This was confirmed with the results of a Pearson's correlation, which showed no significant correlations between subtests of the CCIID and track performance scores. A post-hoc power analysis for this regression analysis revealed an observed power of 0.46 which is below the desired level of 0.80 (Pallant, 2005; Field, 2005).

Consequently, the null hypothesis that there is no association between the scores on the CCIID and sports performance for elite track athletes with ID should be accepted.

4.3.3. Sport-specific study 3: Swimming

The third pilot study included 92 elite swimmers (57 male, 35 female). Table 52 shows minimum and maximum scores, mean and SD for the participants of this study.

Table 52 Descriptive statistics swimming study

	Min.	Max.	Mean	SD	N
age	14	35	22.10	4.67	92
CCIID composite score	-1.29	2.04	.34	.61	61
'Series' scores	3	38	17.38	8.29	63
'Odd One Out' scores	17	41	30.20	4.51	64
'Jigsaw' scores	1	11	6.11	3.21	82
swimming performance scores	1.07	1.66	1.42	0.11	92

All values were within 3.29 SD of the mean and, therefore, no outliers were identified (Tabachnick & Fidell, 2007).

A hierarchical regression showed that there was no significant association between the swimming performance scores and the CCIID scores, or the subtest scores, when controlling for sex and age. This was confirmed with the results of a Pearson's correlation, which showed no significant correlations between subtests and swimming performance scores. A post-hoc power analysis for this regression analysis revealed an observed power of 0.06,

which is below the desired level of 0.80 (Pallant, 2005; Field, 2005). Therefore, the probability to correctly reject a false null hypothesis is unacceptably low. The consequences of the low power level will be considered in the discussion.

Consequently, the null hypothesis that there is no association between the scores on the CCIID and sports performance for elite swimmers with ID should be accepted.

4.4. Discussion

The results of the sport-specific studies confirmed the findings of the initial studies: The relationship between cognitive abilities and sports performance depends on the sports discipline. For table tennis players, results revealed a significant association between cognitive abilities and sports performance. When looking at the subtests separately, the subtest 'Series' was the best predictor for table tennis performance. The subtest 'Series' is a nonverbal test for inductive reasoning abilities and is similar to subtests included in nonverbal intelligence tests (see section 3.1.5). This supports the outcome of the initial studies, which demonstrated that inductive reasoning abilities are related to table tennis performance (see section 2.3).

As discussed in chapter 1, inductive reasoning abilities might be necessary to make tactical decisions, the processing of information on the ball's trajectory and the subsequent planning of motor response (Vickers, 1996). These results support separate sport events for table tennis players with ID.

For track athletes and swimmers, the results showed no association between the scores on the CCIID and sport performance scores. Again, these sport-specific findings are in line with the results of the initial studies. Although the results of the sport-specific studies show a significant association between cognitive abilities and sports performance for table tennis, but not track athletes and swimming, it cannot be definitely concluded that intellectual disability only has an impact on table tennis, but not on track athletics and swimming sports.

Firstly, the CCIID includes only a limited range of subtests (for inductive reasoning and visual processing), and other cognitive abilities, such as memory or attention could be related to sports performance in these disciplines. Secondly, limitations in adaptive behaviour could affects sports performance of individuals with ID, which were not taken into account in the

present study. Limitations in adaptive behaviour are part of the diagnosis of intellectual disability (see chapter 1), but were not part of this thesis. Thirdly, the linear mixed effects models used for track and swimming performance, controlled for distance in order to increase the number of cases included in the analysis. Based on the performance scores, it is, therefore, not possible to examine the performance score for different distances. As swimming and running over longer distances require more pacing strategy (Tucker, Lambert & Noakes, 2006) than for shorter distances, it could be assumed that cognitive abilities have more impact on longer distance runs/swims (e.g. 1500 m swimming or 5000 m track) than short distances (e.g. 100 m swimming or track). However, there were not enough athletes for long distance swimming or track athletics in the 2009 Global Games to look at these distances separately. Further research should investigate the relationship between cognitive abilities and sports performance for long distances separately.

Consequently, it would be necessary to further investigate the relationship between intellectual disability and sport performance for track athletics and swimmers, with a wider range of cognitive tests, appropriate scales for adaptive behaviour and different distances, before definite conclusions can be drawn.

The CCIID proved to be an appropriate tool to assess cognitive abilities of elite athletes with ID. The test battery did not show floor or ceiling effects (only 5 participants had floor or ceiling scores in the subtest 'Jigsaw', none of the other subtests showed ceiling or floor scores), which supports, again, the validity of the test battery for this population. In addition, while comments of participants had indicated culturally biased items in the SON-R intelligence test in the initial studies (see chapter 1.3), no such comments were made for the CCIID by either participants or sports coaches who watched the assessment. This indicates a high degree of cultural fairness for the test battery. Although the assistance of the sport coaches to translate and

communicate the instructions for the subtests was regularly required, the test assistants subsequently felt confident that the participants understood the tasks. The extent to which this confidence was justified should be investigated in future studies.

The sport-specific studies had several limitations. First, there were a number of cases, for which the scores would be rated as unreliable and these were excluded from the analysis: Although sport coaches were present to translate the instructions if necessary, and asked not to interrupt during the testing phase, they would occasionally continue to communicate with the participant after the examples. Also, in some cases, coaches could not be present to explain the instructions in the native language of the participant. In addition, at times interruptions occurred or a participant showed a clear lack of motivation to complete the tests accurately, despite encouragement from the test assistant (and, during the examples, the sports coach). These instances were noted by the test assistant and these data points were subsequently removed from the analysis. For further assessments with the CCIID, it should be ensured that coaches are always present to translate, but are asked to not interfere after the examples. Therefore, coaches should be more clearly informed prior to the start of the assessment by the test assistant. Furthermore, the lack of motivation might have been due to exhaustion of the participant. Although all participants were given ample time to recover if they had competitions on the same day, they still might be affected in their cognitive performance. Further investigations should establish the reasons for the lack of motivation to complete the tests accurately. Nevertheless, all unreliable data was removed prior to the analysis and the findings are, therefore, not affected by these factors.

However, this may affect generalisability of the results as these occurrences may be a normal part of the testing routine.

Second, the relationship between cognitive abilities and sports performance was not investigated for different origins of intellectual disability as the cause of intellectual disability is known only in a minority of cases (Kaski, 2009). The nature of the disability might, however, affect the relationship between cognitive abilities and sports performance: Dellavia, Pallavera, Orlando and Sforza (2009), for example, showed that individuals with Down syndrome had less postural stability than individuals with non-syndromic ID or individuals without ID. Postural stability might influence the performance in certain sport disciplines in which postural stability is an important factor such as gymnastics (Asseman, Caron & Crémieux, 2005). Furthermore, studies showed that very preterm born children or children with very low birth weight have less strength, endurance, flexibility, movement control and hand-eye coordination than their full term and normal weight peers (Sagnol, Debillon & Debû, 2007; Rogers, Fay, Whitfield, Tomlinson & Grunau, 2005). These studies indicate that for some causes of ID the origin of the disability is related to different physical impairments. This might affect the relationship between cognitive abilities and sports performance. However, as the identification of the cause of ID is only possible in a minority of cases, it cannot be included as a factor in the relationship between cognitive abilities and sports performance for athletes with ID.

Third, the relationship between cognitive abilities and sports performance might be influenced by lifestyle factors such as dietary habits, smoking or drinking. These factors were not included in the sport-specific studies, as they are very difficult to assess in training or competition settings as the assessment would have to rely on self-reports. Self – reports, however, are proven to be an unsatisfactory method for individuals with ID, as they show a low consistency of responses (Ruddick & Oliver, 2005). However, all participants included in the sport-specific studies were top-level athletes and, therefore, an overall relatively healthy lifestyle among the majority of participants could perhaps be safely assumed. Further studies should

investigate differences in lifestyle between elite athletes with and without ID to confirm this assumption.

Fourth, an a-priori power analysis showed that the table tennis study had an insufficient number of participants, and retrospective power analyses for the swimming and track athletics studies indicated an unacceptable low power. Commonly, this would affect the generalizability of the results. However, due to the nature of this population (elite athletes with intellectual disabilities), the samples used in the analyses represented a large part of the population. Therefore, in this particular case, it is likely that the results can be generalized to the overall population of elite athletes with intellectual disabilities.

In sum, the results of the sport-specific studies reflect the findings of the initial studies and indicate that the relationship between cognitive abilities and sports performance depends on the sports discipline. While for table tennis players the results demonstrated a clear association between some cognitive abilities and sports performance, this was not the case for swimmers or track athletes. However, as the CCIID contained only a limited number of tests to assess cognitive abilities, and the assessment with a wider range of cognitive tests and an adaptive behaviour scale will be necessary to draw definite conclusions concerning the impact of intellectual disability on sports performance for these disciplines. Nevertheless, the assessment with the CCIID supported the validity and cultural fairness for testing individuals with ID from different cultural backgrounds.

Chapter V: Discussion

This thesis aimed to investigate the association between cognitive abilities and sports performance in elite athletes with ID. A review of literature (see chapter 1: Introduction) indicated that for non-disabled athletes, the relationship between cognitive abilities and sports performance depends on sports discipline. Furthermore, the literature suggested that there is a significant difference in physical performance between individuals with and without ID. However, none of these studies investigated the association between the degree of cognitive functioning and physical performance. Therefore, several intelligence tests were compared to identify the most suitable test to examine the relationship between the degree of cognitive disability and physical performance. The selection was based on a) an analysis of psychometric criteria and b) suitability for the target group (teenagers and young adults with ID from different cultural backgrounds). Subsequently, the relationship between the degree of cognitive abilities and physical performance of athletes with ID was examined using the most suitable intelligence test which was deemed to be the SON-R, and the ABC physical aptitude test. This relationship was explored for all athletes together as well as for the individual sports disciplines of recreational football, track and field athletics, table tennis and swimming. The results, described in chapter 2, confirmed that a significant association was present between scores on the nonverbal SON-R intelligence test and physical performance as measured with the ABC physical aptitude test. Furthermore, when looking at the different sports disciplines separately, the results indicated that this relationship depended on sports discipline: while for elite table tennis and recreational football, inductive reasoning abilities were significantly associated with physical performance and for swimmers, visual processing was related to physical performance, there was no association for track and field athletes between cognitive abilities and physical performance. When using sports

200

competition outcomes, this relationship was confirmed for table tennis, but not for swimming (sports competition outcomes could not be obtained for recreational football players or track and field athletes).

In addition, several test items were found to be culturally sensitive, and floor effects limited the discrimination between different degrees of cognitive functioning towards the lower end of the spectrum. Therefore, although the results confirmed an association in some sports disciplines, such as table tennis players and swimmers, the limitations of the intelligence test showed the need for a new instrument, specifically developed for the target group and the purpose of the study. This led to the development of the Computerized Cognitive test battery for Individuals with Intellectual Disabilities (CCIID).

The design of the CCIID was based on the Cattell-Horn-Carroll theory of cognitive abilities, and included considerations concerning a) the target group (and the need for resolution towards the lower IQ bands), b) cultural fairness, c) results of the initial studies described in chapter 2, and d) a review of designs of different nonverbal subtests. The CCIID was then employed in special schools, colleges and organisations for individuals with ID including 69 participants aged from 15 to 44 years. This initial assessment using modern and classical test theories led to a revision of several items. Psychometric properties of the CCIID were assessed based on investigations of validity and reliability including 91 individuals with ID.

Subsequently, the association between cognitive abilities and sports performance was re-investigated in 218 elite athletes with ID who were participating in international sports competitions using the new test battery. Results confirmed the outcome of the initial studies: the relationship between cognitive abilities and sports performance depended on sports discipline. For table tennis players, sports performance was predicted by the subtest 'Series', which is a nonverbal test for inductive reasoning abilities. Sports performance of swimmers and track athletes, however, was not found to be significantly associated with any of the subtests, or the overall CCIID score.

These results reflected earlier findings for non-disabled athletes, which indicated that the relationship between cognitive abilities and sports performance depends on sport discipline. As discussed in chapter 1, several studies demonstrated that non-disabled athletes participating in team sports, or individual sport disciplines based on interaction with an opponent, performed significantly better on cognitive tasks than athletes participating in individual sports based on speed, such as track and field athletics and swimming (Ryan, Atkinson and Dunham, 2004; Kasahara, Mashiko & Niwa, 2008; Overney, Blanke and Herzog, 2008). This suggests that team sports and sports disciplines primarily based on interaction with an opponent require more cognitive skills, and are consequently more affected by cognitive impairments, than individual sport disciplines based on speed.

Furthermore, the results of this thesis corresponded to some extent with the outcomes of studies investigating the difference between athletes with and without ID, which demonstrated that intellectual disability has a significant impact on physical performance (Van de Vliet & al., 2006; Frey & al, 1999). The studies included in chapter 2 of this thesis showed that physical performance as measured with the ABC physical aptitude test was related to cognitive abilities for table tennis players and swimmers, but not for track and field athletes. When including performance in sports competitions, the results indicated that only for table tennis players sports performance was related to cognitive abilities, but not for swimmers and track athletes. Therefore, the conclusions of the above studies may well be valid, but should be confirmed for different sport disciplines separately.

Nevertheless, although inductive reasoning and visual processing are both key indicators for the general intelligence factor "g" (McGrew, 2005; Snow, Kyllonen & Marshalek, 1984), and are likely to be strongly related to other aspects of intellectual functioning, intellectual disability might affect sports performance through other cognitive aspects, which were not included in the

202

CCIID, such as memory or attention. Further research with tests including tasks for those cognitive abilities should establish if sports performance in swimming and track and field athletics is limited through impairments of other cognitive functions. In addition, another criterion of intellectual disability is 'limitations in adaptive behaviour' (see chapter 1). These limitations could also have an effect on sports performance but were not part of this thesis.

The study had several limitations: As discussed in chapter 1, there are several biological factors that can underlie intellectual disability as well as cause impairments in physical performance. These origins of the intellectual disability, such as very preterm birth, or genetic syndromes (Down syndrome, William's syndrome, fragile x-syndrome) were not considered in the studies of this thesis, although they might be a factor in the relationship between cognitive abilities and physical performance. For example, Dellavia, Pallavera, Orlando and Sforza (2009) showed that individuals with Down syndrome had less postural stability than individuals with non-syndromic ID or individuals without ID. Postural stability might affect sport performance of disciplines where postural stability is an essential factor such as gymnastics (Asseman, Caron & Crémieux, 2005). Furthermore, studies indicated that children who are born very preterm or with very low birth weight have less strength, endurance, flexibility, movement control and hand-eye coordination than their full term and normal weight peers (Sagnol, Debillon & Debû, 2007; Rogers, Fay, Whitfield, Tomlinson & Grunau, 2005). Therefore, it could be concluded that different causes of ID are related to different physical impairments. This might influence the relationship between cognitive abilities and sports performance. However, despite these possible differences, individuals with ID are regularly grouped together for physical exercise (e.g. in special schools) training and competition. Thus, it is important to establish the relationship between cognitive abilities and physical performance for the group of individuals with ID as such. In addition, underlying biological causes

of the intellectual disability, can only be established for a minority of individuals with ID (Kaski, 2009).

A second limitation of the sport performance studies in this thesis is the assumption of a healthy lifestyle and motivation. Most studies investigating the difference in physical performance between individuals with and without ID did not control for lifestyle factors, such as such as dietary habits, smoking and drinking (see chapter 1). These factors are difficult to assess, particularly in a training or competition setting, where the studies of this thesis were conducted. Firstly, the time athletes can dedicate to testing is limited due to constraints of their training and competition schedule. Secondly, as those people who assist the athletes in their daily living (e.g. parents or supervisors), were not accompanying the athletes to trainings or competitions, the evaluation of lifestyle and motivation would have to rely on self-reports, which have proven to be an unsatisfactory method for individuals with ID as they show a low consistency of responses (Ruddick & Oliver, 2005). However, as participants included in the sports performance studies covered by this thesis were elite athletes (with exception of study 1: recreational football players) who took part in national and international sports competitions, the assumption of a healthy lifestyle and a high level of motivation seemed justified. Further research should investigate differences in lifestyle and motivation between elite level athletes with and without ID to confirm this assumption.

The third limitation of this thesis concerns the cross-cultural fairness of the CCIID. Although cultural fairness of the test battery had been considered comprehensively using Athanasiou's framework for test fairness in the construction process, no test is entirely free from cultural bias (Sattler, 1992). However, while items included in the initially used SON-R intelligence test received several comments from participants who did not recognize the objects depicted in the items (see chapter 2), no comments were made from

either participants or their coaches indicating a possible cultural bias of items in the CCIID. In order to ensure that cross-cultural bias is limited in the assessment with the CCIID, the degree of cultural bias should be investigated empirically. Due to practical limitations, this was not possible within the frame of this thesis and should be examined in future studies.

The last limitation concerns methodological issues: The sample size of several studies was rather small (pilot project, studies 1 to 5, sport-specific studies and latent trait model analysis). In several cases, a power analysis indicated an unacceptable low power. Commonly, this would affect the generalizability of the results. However, due to the nature of this population (elite athletes with intellectual disabilities), the samples used in the analyses represented a large part of the population. Therefore, in this particular case, it is likely that the results can be generalized to the overall population of elite athletes with intellectual disabilities.

Power could for most studies only be calculated retrospectively (except for sport-specific study 1: table tennis) as no prior research had established observed effect sizes or variances an a priori power calculation. Therefore, it was not possible to determine the required sample size in the planning stages of the studies, which would have been the preferred method (Thomas, 1997). However, due to practical limitations, it is doubtful if larger samples could have been obtained for these studies.

Although a larger sample for the latent trait model analysis for the subtests 'Series' and 'Odd One Out' would have been desirable, any revisions of the test battery were based on a combination of the results of the latent trait model analysis and the proportion of correctly scored items. In addition, the proportions of correctly scored items were given more weight in the decisions for alterations than the results of the latent trait model analysis as, due to the small sample size, the latent trait model analysis could not obtain the exact misfit of items (Harwell and Janosky, 1991).

205

The results of this thesis contribute to the development of elite sports for athletes with ID in several ways. Firstly, they confirm a clear association between cognitive impairment and sports performance for table tennis players. This association would support separate sport events for individuals with ID. As discussed in chapter 1, separate sport events for individuals with ID are only justified if the athletes cannot reach the same level of performance due to their disability. Furthermore, the confirmation of the impact of the intellectual disability on sport performance is also a pre-requisite for the re-inclusion of athletes with ID into the Paralympics. The results of this thesis showed that, depending on sports discipline, the degree of cognitive functioning, which is part of the criteria for intellectual disability, is indeed associated with sports performance in athletes with ID. Therefore, the results of this thesis support the re-inclusion of athletes with ID in the Paralympics for table tennis. As the findings did not indicate an association between the cognitive abilities included in the CCIID and sports performance for swimming and track athletics, further research using a wider range of cognitive tests should establish the association between other cognitive functions, such as memory and attention, and sports performance, as well as between adaptive behaviour and sports performance, as adaptive behaviour is also part of the criteria for intellectual disability.

Secondly, the current results could support the development of a sport talent identification system for individuals with ID. Based on the findings, individuals with those cognitive abilities related to superior sports performance in table tennis could be identified. Further research should establish if these cognitive abilities are indeed predictive for future sports performances and to what extent talent identification encourages young individuals with ID to engage in a recommended sport discipline.

This thesis developed a novel computerized cognitive test battery to assess the degree of intellectual functioning in individuals with ID. Several steps have been undertaken to ensure the quality of the test battery. Item difficulty was assessed using modern and classical test theories and items were revised accordingly. The evaluation of psychometric properties confirmed overall reliability and validity of the CCIID. Validity was investigated in an analysis of content, construct and criterion validity. The results of these studies supported the validity of the CCIID as an assessment tool to investigate intellectual functioning of individuals with ID. In addition, detailed examinations of internal consistency, test-retest and inter-rater reliability indicated that the CCIID can be rated as a highly reliable instrument to assess individuals with ID.

The CCIID was developed for the assessment of cognitive abilities of individuals with ID as the initial studies, which were based on the assessment with a nonverbal intelligence test, had indicated the need for an assessment tool specifically developed for individuals with ID from different cultural backgrounds. The CCIID has the potential to be used in several contexts. Although the CCIID was developed as a research tool, it might also be suitable as part of the diagnostic process to identify impairments of nonverbal intellectual functioning. Future studies would have to establish if the CCIID would be suitable as a tool to discriminate between verbal and nonverbal cognitive impairments, and if it could be used as an instrument to identify developmental deficits of inductive reasoning and visual processing in a school context.

In sum, this thesis investigated the relationship between cognitive abilities and sports performance in elite athletes with ID and developed a computerized cognitive test battery for individuals with ID from diverse cultural backgrounds. Results showed that there is a significant association between cognitive abilities and sports performance for elite table tennis players, but not

for swimmers or track athletes. These results clearly demonstrated the impact of cognitive impairments on table tennis performance and provided, therefore, a starting point for the re-inclusion of table tennis in the Paralympics. Further research should investigate if sports performance of swimmers and track and field athletes is limited through different cognitive abilities or adaptive behaviour.

Additionally, the assessment of psychometric properties of the CCIID demonstrated that the test battery is a valid and reliable instrument to assess cognitive functioning in individuals with ID. The CCIID proved to be a very suitable tool for the assessment of athletes with ID during sports competitions, but could also be used in different contexts.

References

Aiken, L. R. (1994). *Psychological testing and assessment (8th ed.).* Boston, Massachusetts: Allyn and Bacon.

Alario, F. X. & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability and age of acquisition. *Behaviour Research Methods, Instruments and Computers, 31*(3), 531-552.

Alfonso, V. C., Flanagan, D. P. & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests and issues* (pp. 185-202). New York: The Guilford Press.

Allin, M., Matsumoto, H., Santhouse, A. M., Nosarti, C., AlAsady, M. H. S., Stewart, A. L., Rifkin, L. & Murray, R. M. (2001). Cognitive and motor function and the size of the cerebellum in adolescents born very pre-term. *Brain, 124*(1), 60-66.

Allison, M. G. & Ayllon, T. (1980). Behavioral coaching in the development of skills in football, gymnastics and tennis. *Journal of Applied Behavioral Analysis, 13*(2), 297-314.

American Educational Research Association (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Asseman, F., Caron, O. & Crémieux, J. (2005). Effects of the removal of vision on body sway during different postures in elite gymnasts. *International Journal of Sports Medicine, 26*(2), 116-119.

Athanasiou, M. S. (2000). Current nonverbal assessment instruments: A comparison of psychometric integrity and test fairness. *Journal of Psychoeducational Assessment*, 18, 211-229.

Bartram, D. (1990). Reliability and validity. In J. R. Beech & L. Harding: *Testing people: A practical guide to psychometrics*. Windsor, England: NFER-Nelson Publishing Company Ltd.

Baumgartner, T. A. & Horvath, M. A. (1988). Problems in measuring the physical and motor performance of the handicapped. *Journal of Physical Education, Recreation and Dance, 59*, 48-52.

Belligni, E. F., Biamino, E., Molinatto, C., Messa, J., Pierlluigi, M., Faravelli, F., Zuffardi, O., Ferrero, G. B. & Silengo, M. C. (2009). Subtelomeric FISH analysis in 76 patients with syndromic developmental delay/intellectual disability. *Italian Journal of Paediatrics, 35*(9). Retrieved August 9, 2009, from http://www.ijponline.net/content/35/1/9.

Black, D. P., Smith, B. A., Wu, J. & Ulrich, B. D. (2007). Uncontrolled manifold analysis of segmental angle variability during walking: Preadolescents with and without Down syndrome. *Experimental Brain Research, 183*(4), 511-521.

Boles, D. B. & Law, M. B. (1998). A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 204-215.

Braden, J. P. & Athanasiou, M. S. (2005). A comparative review of nonverbal measures of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests and issues* (pp. 557-578). New York: The Guilford Press.

Carmeli, E., Bar-Yossef, T., Ariav, C., Levy, R. & Liebermann, D. G. (2008). Perceptual-motor coordination in persons with mild intellectual disability. *Disability Rehabilitation, 30*(5), 323-329.

Carmeli, E., Bar-Yossef, T., Ariav, C., Paz, C., Sabbag, H. & Levy, R. (2008). Sensorimotor impairment and strategies in adults with intellectual disabilities. *Motor Control, 12*(4), 348-361.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cattell R. (1949). *Test of "g": Culture fair.* Savoy (IL): Institute for Personality and Ability Testing.

Cattell, R. (1971). *Abilities, their growth, structure and action.* Boston: Houghton Mifflin.

Charlton, J. L., Ihsen, E. & Lavelle, B. M. (2000). Control of manual skills in children with Down syndrome. In D. J. Weeks, R. Chua & D. Elliott (Eds.), *Perceptual motor behavior in Down syndrome* (pp. 25-48). Champaign, IL: Human Kinetics.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, L. G. & Spenciner, L. J. (2003). *Assessment of children and youth with special needs* (2nd ed.). Boston: Pearson Education.

Cooper, S. A., Melville, C. & Morrison, J. (2004). People with intellectual disabilities. *British Medical Journal, 329*(7463), 414-415.

Cornwell, K. L. (2004). People with intellectual disabilities – people registered disabled with learning disabilities tend to fall through the net. *British Medical Journal, 329*(7471), 917.

Cressler, R. V., Lavay, B. & Giese, M. (1988). The reliability of four measures of cardiovascular fitness with mentally retarded adults. *Adapted Physical Activity Quarterly, 5,* 285-292.

Croce, R. V. (1990). The effects of exercise and diet on body-composition and cardiovascular fitness in adults with severe mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 25*(2), 176-187.

Croce, R. V., Horvat, M. & Pitetti, K. H. (1999). Prediction of dynamic strength in individuals with mental retardation. *Clinical Kinesiology, 52,* 79-86.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th ed.). New York: Harper & Row.

Dacey, C. M., Nelson, W. M. 3rd & Stockel, J. (1999). Reliability, criterion-related validity and qualitative comments of the Fourth Edition of the Stanford-Binet Intelligence Scale with a young adult population with

intellectual disability. *Journal of Intellectual Disability Research, 43*(3), 179-184.

Davidson, J. E. & Downing, C. L. (2000). Contemporary models of intelligence. In R.J. Sternberg (Ed.), *Handbook of intelligence* (pp.34-49). New York: Cambridge University Press.

Davis, P. & Sime, W. (2005). Towards a psychophysiology of performance: Sport psychology principles dealing with anxiety. *International Journal of Stress Management*, 12(4), 393-378.

De Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: The Guilford Press.

De Koning, E., Sijtsma, K. & Hamers, J. H. M. (2003). Construction and validation of a test for inductive reasoning. *European Journal of Psychological Assessment. 19*(1), 24-39.

Dellavia, C., Pallavera, A., Orlando, F. & Sforza, C. (2009). Postural stability of athletes in Special Olympics. *Perceptual and Motor Skills, 108*(2), 608-622.

Del Villar, F., Gonzalez, L. G., Iglesias, D., Moreno, M. & Cervello, E. M. (2007). Expert-novice differences in cognitive and execution skills during tennis competitions. *Perceptual and Motor Skills, 104*(2), 355-365.

Dinnage, R. (1986). *The child with cerebral palsy.* Great Britain: N-FER-Nelson Publishing Company Ltd.

Draheim, C. C., Williams, D. P. & McCubbin, J. A. (2002). Prevalence of physical inactivity and recommended physical activity in community-based adults with mental retardation. *Mental Retardation, 40*(6), 436-444.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96* (3), 433-458.

Eichstaedt, C. B., & Lavay, B. W. (1992). *Physical activity for individuals with mental retardation: Infant through adulthood.* Champaign, IL: Human Kinetics.

Ericsson, K. A. & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual review of Psychology, 47*, 273-305.

Fait, H. & Dunn, J. (1984). *Special physical education: Adapted, individual and developmental* (5th ed.). Philadelphia: W.B. Sauders.

Fernhall, B. (1993). Physical fitness and exercise training of individuals with mental retardation. *Medicine and Science in Sports and Exercise, 25*, 442-450.

Fernhall, B., Millar, L. A., Tymeson, G. & Burkett, L. N. (1991). Maximal exercise testing of mentally retarded adolescents and adults: reliability study. *Archives of Physical Medicine and Rehabilitation, 71*, 1065-1068.

Fernhall, B., Pitetti, K. H., Rimmer, J. H., McCubbin, J. A., Rintala, P. Millar, A. L., Kittredge, J. & Burkett, L. N. (1996). Cardiorespiratory capacity of individuals with mental retardation including Down syndrome. *Medicine and Science in Sports and Exercise, 28*(3) 366-371.

Fernhall, B., Pitetti, K. H., Vukovich, M. D., Stubbs, N., Hensen, T., Winnick, J. P. & Short, F. X. (1998). Validation of cardiovascular fitness field tests in children with mental retardation. *American Journal on Mental Retardation, 102*, 602-610.

Fernhall, B., Tymeson, G. T. & Webster, G. E. (1988). Cardiovascular fitness of mentally retarded individuals. *Adapted Physical Activity Quarterly, 5*, 12-28.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London, UK: Sage Press.

Flanagan, D. P. & Harrison, P. L. (Eds.). (2005). *Contemporary intellectual assessment. Theories, tests and issues*. New York: The Guilford Press.

Flanagan, D. P. & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment.* Hoboken, NJ: John Wiley & Son.

Forsythe, A., Mulhern, G. & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods, 40*(1), 116-129.

Fox, R, Burkhart, J. E. & Rotatori, A. F. (1983). Appropriate classification of obesity of mentally retarded adults. *American Journal of Mental Deficiency, 88*, 112-114.

Frey, G. C., McCubbin, J. A., Hannigan-Downs, S., Kassers, S. L. & Skaggs, S. O. (1999). Physical fitness of trained runners with and without mild mental retardation. *Adapted Physical Activity Quarterly, 16*, 126-137.

Ghilelli, E. E., Campbell, J. P. & Zedek, S. (1981). *Measurement theory for the behavioral sciences*. New York: W.H. Freeman.

Gunning-Dixon, F. M. & Raz, N. (2000). The cognitive correlates of white matter abnormalities in normal aging: A quantitative review. *Neuropsychology, 14*, 224-232.

Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence: Vol. 4.* (pp. 35-71), New Jersey: Erlbaum.

Hagerman, R. J. & Hagerman, P. J. (2002). The fragile X permutation: into the phenotypic fold. *Current Opinion in Genetics and Development, 12*(3), 278-283.

Halle, J. W., Silverman, N. A. & Regan, L. (1983). The effects of a data-based exercise program on physical fitness of retarded children. *Education and Training in Mental Retardation, 18*, 221-225.

Hall, A., McLeod, A., Counsell, C., Thomson, L. & Mutch, L. (1995). School attainment, cognitive ability and motor function in a total Scottish very-low-birth weight population at eight years: A controlled study. *Developmental Medicine and Child Neurology, 37*, 1037-50.

Hambleton, R. K. (1991). *Fundamentals of Item Response Theory*. California,Newbury Park: Sage Publications.

Hambleton, R. K. & Rovinelli, R. J. (1973). A Fortran IV program for generating examinee response data from logistic test models. *Behavioral Science, 17*, 73-74.

Hammill, D. D., Pearson, N. A. & Wiederholt, J. L. (1997). *Comprehensive Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.

Harris, A. M., Reynolds, M. A. & Koegel, H. M. (1996). Nonverbal assessment: Multicultural perspectives. In L. A. Suzuki, P. J. Miller & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological and educational applications.* (pp.223-252). San Francisco: Jossey-Bass.

Harrison, T. B. & Stiles, J. (2008). Hierarchical forms of processing in adults and children [Electronic version]. *Journal of Experimental Child Psychology, 103*(2), 222-240.

Harwell, M. R. & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied psychological measurement, 15*, 279-291.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T. & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *The Journal of Applied Psychology, 92*(2), 373-385.

Hill, C., Croce, R., Miller, J. & Cleland, F. (1996). Muscle torque relationships between hand-held dynamometry and isokinetic measurements in children ages 9-11. *Journal of Strength Conditioning Research, 10,* 77-82.

Hodges, N. J. & Franks, I. M. (2002). Modelling coaching practice: the role of instruction and demonstration. *Journal of Sports Sciences, 20*(10), 793-811.

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., Mbonani, L. & Stevenson, J. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenia. *Journal of the international neuropsychological society, 10*(2), 246-260.

Horvat, M. (1990). *Physical education and sport for exceptional students.* Dubuque: W. C. Brown Publishers.

Hove, O. (2004). Weight survey on adult persons with mental retardation living in the community. *Research in Developmental Disabilities, 25,* 9-17.

Howell, D.C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove/CA: Duxbury.

215

Hutt, M. L. (1932). The Kohs Block-designs test: A revision for clinical practice. *Journal of Applied Psychology, 16*, 298-307.

International Paralympic Committee (2007). Memorandum. Retrieved November 5, 2008, from http://www.paralympic.org/release/Main_Sections_Menu/Classification/2006_12_19_INAS_Position_Statement.pdf

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrica, 39*, 31-36.

Kane, J. E. (1979). Cognitive aspects of performance. *British Journal of Sports Medicine, 12*, 101-107.

Kasahara, S., Mashiko, H. & Niwa, S. I. (2008). Superior performance in WAIS-R block design among top-level rugby players. *British Journal of Sports Medicine, 42*(11), 632-633.

Kaski, M. (2009). Aetiology of intellectual disability: general issues and prevention. In M.G. Gelder, N.C. Andreasen, J.J. Lopez-Ibor Jr. & J.R. Geddes (Eds.), *New Oxford Testbook of Psychiatry – Second Edition* (pp. 1830-1877). Oxford: Oxford University Press.

Kaufman, A. S. & Kaufman N. L.(1983). *K-ABC interpretive manual.* Circle Pines, MN: American Guidance Service.

Kaufman, J. C. (2003). Gender and ethnic differences on the KABC-II. In M.H. Daniel (Ed.), *The KABC-II: Theory, content and administration. Symposium presented at the annual meeting of the American Psychological Association.* Toronto.

Kaufman, A. S. & Kaufman N. L. (2004). *Manual for the Kaufman Assessment Battery for Children - Second Edition (KABC-II), comprehensive form.* Circle Pines, MN: American Guidance Service.

Kaufman, J. C., Kaufman A. S., Kaufman-Singer, J. & Kaufman, N. L. (2005). The Kaufman Assessment Battery for Children – Second Edition and the Kaufman Adolescent and Adult Intelligence Test. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests and issues* (pp. 344-370). New York: The Guilford Press.

Keith, T. Z., Goldenring Fine, J., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Hierarchical, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review, 35,* 108-127.

Klanderman, J., Devine, J. & Mollner, C. (2006). The K-ABC: A construct validity study with the WISC-R and Stanford-Binet. *Journal of Clinical Psychology,* 41(2), 273-281.

Kline, P. (1993). *The Handbook of psychological testing.* New York: Routledge.

Kline, Th. J. B. (2005). *Psychological testing. A practical approach to design and evaluation.* Thousand Oaks/California: Sage Publications.

Kubo, M. & Ulrich, B. (2006). Coordination of pelvis-HAT (head, arms and trunk) in anterior-posterior and medio-lateral directions during treadmill ait in preadolescents with/without Down syndrome. *Gait Posture, 23*(4), 512-518.

Lahtinen, U., Rintala, P. & Malin, A. (2007). Physical performance in individuals with intellectual disabilities: A 30-year follow-up. *Adapted Physical Activity Quarterly, 24*(2), 125-143.

Lavay, B., Reid, G. & Cressler-Chaviz, M. (1990). Measuring the cardiovascular endurance of persons with mental retardation: A critical review. *Exercise and Sport Science Review, 18,* 263-290.

Leger, L. A., Mercier, D., Gadoury, C. & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Science, 6,* 93-101.

Lichtenberger, E. O. (2004). Autistic-spectrum disorders. In D. P. Flanagan & A. S. Kaufman (Eds.), *Essentials of WISC-IV assessment.* New Jersey: John Wiley & Sons.

Lohman, D. F. (1994). Spatially gifted, verbally inconvenienced. In N. Colangelo, S. G. Assouline & D. L. Ambroson (Eds.), *Talent development: Vol. 2.* Proceedings from the 1993 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development (pp.251-340). Dayton, OH: Ohio Psychology Press.

Lopez, E. C. (1997). The cognitive assessment of limited English proficient and bilingual children. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.) *Contemporary intellectual assessment: Theories, tests and issues* (pp. 32-47). New York: Guilford Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Mac Donncha, C, Watson, A. W. S., McSweeny, T. & O'Donnovan, D. J (1999). Reliability of Eurofit physical fitness items for adolescent males with and without mental retardation. *Adapted Physical Activity Quarterly, 16,* 86 – 95.

Machek, M.A., Stopka, C.B., Tillman, M.D., Sneed, S.M. & Naugle, K.E. (2008). The effects of a supervised resistance-training program on Special Olympics athletes. *Journal of Sport Rehabilitation, 17*(4), 372-379.

Maksud, M. G. & Hamilton, L. H. (1975). Physiological responses of EMR children in strenuous exercise. *American Journal of Mental Deficiency, 79,* 32-38.

Mann, D. T. Y., Williams, A. M. Ward, P. & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology, 29*(4), 457-478.

Marois, R. & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences, 9*(6), 296-305.

Marshalek, B., Lohman, D. F. & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107-128.

Mayer, R. E (1994). Reasoning, Inductive. In R.J . Sternberg: *Encyclopedia of human intelligence* (Vol. 2, pp. 935-938). New York: Macmillan Publishing Company.

McCallum, R. S. (2003). Context for nonverbal assessment of intelligence and related abilities. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 3-22). New York: Kluwer Academic/Plenum Publishers.

McGrew, K. S., Werder, J. K. & Woodcock, R. W. (1991). *WJ-R technical manual*. Chicago: Riverside.

McGrew, K. S., Werder, J. K. & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Rverside.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual Assessment. Theories, tests and issues* (pp. 136-181). New York: The Guilford Press.

Miller, R. J. (1973). Cross-cultural research in the perception of pictorial materials. *Psychological Bulletin, 80*, 135–150.

Mondloch, C. J., Geldart, S., Maurer, D & de Schonen, S (2003). Developmental changes in the processing of hierarchical shapes continues into adolescence. *Journal of Experimental Child Psychology, 84* (1), 20-40.

Morris, C. A. & Mervis, C. B. (1999). Williams syndrome. In S. Goldstein & C. R. Reynolds (Eds.), *Handbook of neurodevelopmental and generic disorders in children* (pp. 555-590). New York: Guilford Press.

Mouelhi Guizani, S., Tenenbaum, G., Bouzaouach, I., Ben Khelder, A., Feki, Y. & Bouaziz, M. (2006). Information-processing under incremental levels of physical loads: Comparing racquet to combat sports. *Journal of Sports Medicine and Physical Fitness, 46*(2), 335-343.

Nenty, H. J. & Dinero, T. E. (1981). A cross-cultural analysis of the fairness of the Cattell Culture Fair Intelligence Test using the Rasch model. *Applied Psychological Measurement, 5*, 355-368.

Nieuwenhuys, M. (1991). *A Comparative Study of SON-R, WISC-R en Raven-SPM*. Amsterdam: Internal Report, University of Amsterdam, Department of Developmental Psychology.

Nougier, V., Stein, J. F. & Bonnel, A. M. (1991). Information-processing in sport and orienting of attention. *International Journal of Sport Psychology, 22*(3-4), 307-327.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill Inc.

Nguyen The Tich, S., Anderson, P. J., Shimony, J. S., Hunt, R. W., Doyle, L. W. & Inder, T. E. (2009). A novel quantitative simple brain metric using MR imaging for preterm infants. *American Journal of Neuroradiology, 30*(1), 125-131.

O'Connell, J.K., Rutland, M.K. & O'Connell, D.G. (2006). Grip strength of Texas Special Olympians. *Perceptual and Motor Skills, 102*(2), 461-466.

Oller, J. W., Kim, K. & Choe, Y. (2001). Can instructions to nonverbal IQ tests be given in pantomime? Additional applications of a general theory of signs. *Semiotica, 133* (1-4), 15-44.

Overney, L. S., Blanke, O. & Herzog, M. H. (2008, June 11). Enhanced temporal but not attentional processing in expert tennis players. PLoS ONE, 3(6). Retrieved September 3, 2008, from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002380#pone-0002380-g001

Osuji, O. N. (1982). Constructing complex geometric patterns. *Journal of Cross-Cultural Psychology, 13* (4), 481-489.

Pallant, J. (2005). SPSS survival manual (2nd ed.). New York: Open University Press.

Pearson, N. (2003). Comprehensive Test of Nonverbal Intelligence. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 141-162). New York: Kluwer Academic/Plenum Publishers.

Pena, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*(4), 1255-1264.

Pitetti, K. H. & Campell, K. D. (1991). Mentally retarded individuals – a population at risk? *Medicine and Science in Sports and Exercise, 23*, 586-593.

Pitetti, K. H. & Fernhall, B. (2004). Comparing run performance of adolescents with mental retardation, with and without Down Syndrome. *Adapted Physical Activity Quarterly, 21*(3), 219-228.

Pitetti, K. H. & Tan, D. M. (1990). Cardiorespiratory responses of mentally retarded adults to air-brake ergometry and treadmill exercise. *Archives of Physical Medicine and Rehabilitation, 71*(5), 318-321.

Pitetti, K. H. & Tan, D. M. (1991). Effects of a minimally supervised exercise program for mentally retarded adults. *Medicine and Science in Sports and Exercise, 23*(5), 594-601.

Pitetti, K. H. & Yarmer, D. A. (2002). Lower body strength of children and adolescents with and without mild mental retardation: A comparison. *Adapted Physical Activity Quarterly, 19*, 68-81.

Pitetti, K. H., Yarmer, D. A. & Fernhall, B. (2001). Cardiovascular fitness and body composition of youth with and without mental retardation. *Adapted Physical Activity Quarterly, 18*(2), 127-141.

Pizarro, P. (1990). Reliability of health related fitness test for mainstreamed educable and trainable mentally handicapped adolescents. *Adapted Physical Activity Quarterly, 7*, 240-248.

Powers, S., Jones P. B. & Barkan, J. H. (1986). Validity of the Standard Progressive Matrices as a Predictor of Achievement of Sixth and Seventh Grade Students. *Educational and Psychological Measurement, 46*, 719-722.

The Psychological Corporation (2003). *WISC-IV technical and interpretive manual*. San Antonio, TX: Author.

Rarick, G. L., Widdop, J. H. & Broadhead, G. D. (1970). The physical fitness and motor performance of educable mentally retarded children. *Exceptional Children, 36*, 504-519.

Raven, J., Raven, J. C. & Court, H. J. (1998). *Coloured Progressive Matrices*. Great Britain: Oxford Psychologist Press.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General overview*. San Antonio, TX: Harcourt Assessment.

Reid, G., Montgomery, D. & Seidl, C. (1985). Performance of mentally retarded adults on the Canadian Standardized Test of Fitness. *Canadian Journal of Public Health, 76*, 187-190.

Reise, S. P. & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27,* 133-144.

Rizopoulos, D. (2006) ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25.

Robinson, M. N., Peake, L. J., Ditchfield, M. R., Reid, S. M., Lanigan, A. & Reddhough, D. S. (2008). Magnetic resonance imaging findings in a population-based cohort of children with cerebral palsy. *Developmental Medicine and Child Neurology, 51,* 39-45.

Rogers, M., Fay, T. B., Whitfield, M. F., Tomlinson, J. & Grunau, R. E. (2005). Aerobic capacity, strength, flexibility, and activity level in unimpaired extremely low birth weight (<or=800g) survivors at 17 years of age compared with term-born control subjects. *Pediatrics,* 116(1), 58-65.

Roid, G. H. (2003 a). *Stanford-Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside.

Roid, G. H. (2003 b). *Stanford-Binet Intelligence Scales, Fifth Edition: Technical manual.* Itasca, IL: Riverside.

Roid, G. H. & Miller, L. J. (1997). *Leiter/Stoelting brief nonverbal Intelligence Scale Manual (S-BIT).* Wood Dale, IL: Stoelting.

Roid, G. H. & Miller, L. J. (1999). *Leiter International Performance Scale-Revised (Leiter R) Manual.* Wood Dale, IL: Stoelting.

Roid, G. H., Nellis, L. & McLellan, M. (2003). Assessment with the Leiter International Performance Scale-Revised and the S-BIT. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 113-140). New York: Kluwer Academic/Plenum Publishers.

Roid, G. H. & Pomplun, M. (2005). Interpreting the Stanford-Binet Intelligence Scales, Fifth Edition. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests and issues* (pp. 325-343). New York: The Guilford Press.

Roid, G. H. & Woodcock, R. W. (2000). Uses of Rasch scaling in the measurement of cognitive development and growth. *Journal of Outcome Measurement, 4*, 579-594.

Rösler, F., Heil, M., Bajric, J., Pauls, A. C. & Henninghausen, E. (1995). Patterns of cerebral activation while mental images are rotated and changed in size. *Psychophysiology, 32*(2), 135-149.

Rossili, M. & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition, 52*, 326 -333.

Royer, F. L., Gilmore, G. C. & Gruhn, J. J. (1984). Stimulus parameters that produce age differences in block design performance. *Journal of Clinical Psychology, 40*, 1474 –1485.

Ruddick, L. & Oliver, C. (2005). The development of a health status measure for self-report by people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 18*(2), 143-150.

Rust, J. & Golombok, S. (1989). *Modern Psychometrics: The science of psychological assessment* (2nd ed.). London: Routledge.

Ryan, J. P., Atkinson, T. M. & Dunham, K. T. (2004). Sports-related and gender differences on neuropsychological measures of frontal lobe functioning. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine, 14*(1), 18-24

Sachdev, P. S., Wen, W., Christensen, H. & Jorm, A. F. (2005). White matter hyperintensities are related to physical disability and poor motor function. *Journal of Neurology, Neurosurgery and Psychiatry, 76*(3), 362-367.

Sagnol, C., Debillon, T. & Debû, B. (2007). Assessment of motor control using kinematics analysis in preschool children born very preterm. *Developmental Psychobiology, 49*(4), 421-432.

Sattler, J. M. (1992). *Assessment of Children.* (3rd ed.), San Diego, CA: Author.

Seguin, E. (1907). *Idiocy and its treatment by the psychological method.* New York: Teachers College, Columbia University.

Shim,J., Carlton, L. G., Chow, J. W. & Chae, W. S. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of Motor Behaviour, 37*(2), 164-175.

Skowronski, W., Horvat, M., Nocera, J., Roswal, G. & Croce, R. (2009). Eurofit Special: European fitness battery score variation among individuals with intellectual disabilities. *Adapted Physical Activity Quarterly, 26*(1), 54-67.

Snijders, J. Th., Tellegen, P. J. & Laros, J. A. (1989). *Snijders-Oomen non-verbal Intelligence Test: SON-R 5½ -17. Manual and research report.* Groningen: Wolters-Noordhoff.

Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memeory, 6* (2), 174-215.

Snow, R. E., Kyllonen, P. C. & Marshalek, B. (1984). The topography of ability and learning correlations. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol.2; pp 47-103). Hillsdale, N.J.: Erlbaum.

Steffens, D. C, Bosworth, H. B., Provenzale, J. M. & MacFall, J. R. (2002). Subcortical white matter lesions and functional impairment in geriatric depression. *Depression and Anxiety, 15*(1), 23-28.

Stone, C. A. (1992). Recovery of marginal maimum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement, 16,* 1-16.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistic* (5th ed.). Boston, M.A.: Pearson Education, Inc.

Tartaglia, N., Davis, S., Hench, A., Nimishakavi, S., Beauregard, R., Reynolds, A., Fenton, L., Albrecht, L., Ross, J., Visootsak, J., Hansen, R. & Hagerman, R. (2008). A new look at XXYY syndrome: Medical and psychological features. *American Journal of Medical Genetics, Part A, 146A*(12), 1509-1522.

Tellegen, P. & Laros, J. (1993). The construction and validation of a nonverbal test of intelligence: the revision of the Snijders-Oomen tests. *European Journal of Psychological Assessment, 9*(2), 147-157.

Tellegen, P. & Laros, J. (2004). Cultural bias in the SON-R test: comparative study of Brazilian and Dutch children. *Psicologia: Teoria e Pesquisa, 20*(2), 103-111.

Thomas, K. T. (1994). The development of sport expertise: from Leeds to MVP legend. *Quest, 46*(2), 199-210.

Thomas, L. (1997). Retrospective power analysis. *Conservation Biology, 11*(1), 276-280.

Tomporowski, P. D. & Ellis, N. R. (1985). The effects of exercise on the health, intelligence, and adaptive behavior of institutionalized severely and profoundly mentally retarded adults: a systematic replication. *Applied Research in Mental Retardation, 6*(4), 465-473.

Treisman, A. (1991). Search, similarity and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 652 – 676.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15 – 48.

Tsimaras, V. K. & Fotiadou, E. G. (2004). Effect of training on the muscle strength and dynamic balance ability of adults with Down syndrome. *Journal of Strength and Conditioning Research, 18*(2), 343-347.

Tucker, R., Lambert, M. I. & Noakes, T. D. (2006). An analysis of pacing strategies during men's world record performances in track athletics. *International Journal of Sports Physiology and Performance, 1*(3), 233-245.

Van de Louw, J., Vorstenbosch, R., Vinck, L., Penning, C. & Evenhuis, H. (2009). Prevalence of hypertension in adults with intellectual disabilities in the Netherlands. *Journal of Intellectual Disability Research, 53*(1), 78-84.

Van de Vliet, P., Rintala, P., Frojd, K., Verellen, J., Van Houtte, S., Daly, D. J. & Vanlandewijk, Y. C. (2006). Physical fitness profile of elite athletes with

intellectual disability. *Scandinavian Journal of Science & Medicine in Sports*, 16(6), 417-425.

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology, Human Perception and Performance. 22*, 342-54.

Vinter, A. & Detable, C. (2008). Implicit and explicit motor learning in children with and without Down's syndrome. *British Journal of Developmental Psychology, 24*(4), 507-523.

Wallace, R. A. & Schluter, P. (2008). Audit of cardiovascular disease risk factors among supported adults with intellectual disability attending an aging clinic. *Journal of Intellectual and Developmental Disability, 33*(1), 48-58.

Widjaja, E., Nilsson, D., Blaser, S. & Raybaud, C. (2008). White matter abnormalities in children with idiopathic developmental delay. *Acta Radiologica, 49*(5), 589-595.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio, TX: Psychological Cooperation.

Wechlser, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

World Health Organisation (2007). International Classification of Diseases, Chapter V. Retrieved June 3, 2009, from http://apps.who.int/classifications/apps/icd/icd10online/.

Wuang, Y.-P., Wang, C.-C., Huang, M.-H. & Su, C.-Y. (2008). Profiles and cognitive predictors of motor functions among early school-age children with mild mental disabilities. *Journal of Intellectual Disability Research, 52*(12), 1048-1060.

Appendix A : Information letter and consent form

Dr. Stephan Bandelow
19/6/2008
Department of Human Sciences
Loughborough University
LE11 3TU
email: S.Bandelow@lboro.ac.uk
phone: 01509 223009

Participation in a developmental research project:
Participant information form

Name of project:
Cognitive Testing and Sports and Learning Performance in Learning Disability

Aim of the project:
To assist in the development of reliable test procedures for the classification of athletes who have a learning disability in competitive sports, and to develop test procedures for sensitive measurement of learning performance related to academic (school) progress.

Principal investigators:
Dr Stephan Bandelow <S.Bandelow@lboro.ac.uk>, Dept. of Human Sciences, Loughborough University, LE11 3TU. Phone 01509 223009.
Prof Eef Hogervorst <E.Hogervorst@lboro.ac.uk>, Department of Human Sciences, Loughborough University, LE11 3TU. Phone 01509 223020.

Other relevant information:
The project is supported by the MENCAP sports programme, the Youth Sport Trust, the Bailey Thomas charitable foundation and the Linkage Education Centre for people with learning difficulties. It is also supported by the International Paralympic Committee (IPC) – Intellectual Disability, Exercise and Active Living Research Group (IDEAL-RG). Confidentiality will be ensured by anonymising all collected data, i.e. removing all personal identifiers such as name and address from the data records. Personal data will be kept separate from the test results.

Procedure:
Participants in this project will take part in the following test procedure:

> Several neuropsychological tests to assess brain function. These tests will look at several specific skills, such as attention, eye-hand co-ordination, visual (seeing) speed, reaction times and several types of memory.

A non-verbal IQ-based test, which is designed on the basis of several existing IQ tests.

All tests are computer-based using a touch screen interface. The assessment will take about 25 minutes.

Participation in a developmental research project:
Informed Consent Form

Please complete the following short form if you are willing to take part in this project.

I am willing to take part in this above research project   _____ (tick)

I understand that I can withdraw at any time by informing the   _____ (tick)
organisers

I agree for my test scores to be included in the research data   _____ (tick)

I agree that results may be published in report of academic   _____ (tick)
paper form

HOWEVER – I understand that these scores, and published
results, will not be attributed to me by name   _____ (tick)

I understand that any notes / results held by the investigators
will be destroyed after the appropriate time in accordance   _____ (tick)
with the Data Protection Act

I understand that I can withdraw from this study at any time,
without the need to give any specific reasons for doing so.   _____ (tick)

Date: _____

Participant name: _____

Signature: _____

Parent/Carer name (if appropriate): _____

Signature: _____

LOUGHBOROUGH UNIVERSITY
ETHICAL ADVISORY SUB-COMMITTEE

RESEARCH PROPOSAL
INVOLVING HUMAN PARTICIPANTS

Title: Contributions of IQ and Neuropsychological testing to school and sports performance.

Applicant: Dr S Bandelow

Department: Human Sciences

Date of clearance: 7 December 2007

Comments of the Sub-Committee:

The Sub-Committee agreed to issue clearance to proceed subject to the following conditions:
- That confirmation was provided as to whether a paper-based IQ test would be a scientifically sound way of assessing intelligence in those with learning disabilities
- That Criminal Records Bureau Checks were completed for all investigators for this study
- That the start date for data collection was confirmed. It appeared that data collection had already started and the Sub-Committee emphasized that Ethical Clearance should have been sought prior to the start of data collection
- That a Participant Information Sheet was submitted to the Committee including
  o Full contact details of all investigators
  o In depth details of the procedures, in terminology which would be easily understood by the participants
  o Any exclusion criteria
  o Details of any risks involved and the protocols for addressing those risks
  o Information about data storage, anonymity and compliance with the Data Protection Act
  o A statement regarding the participants right to withdraw from the study at any point with out needing to provide a reason for doing so.

- That additional information was submitted on how investigators would insure that Informed Consent had been obtained, and that the permission form be re-titled 'Informed Consent Form'
- That confirmation of Head of Department approval was provided (an email would suffice).
- That further, in depth details were given on the precise nature of the investigations. This would include details of the physical activities involved.
- Confirmation of University Insurance Cover is provided
- That a Health Screen Questionnaire was completed by all participants

30/01/2008

Subtest 'Series' example 1
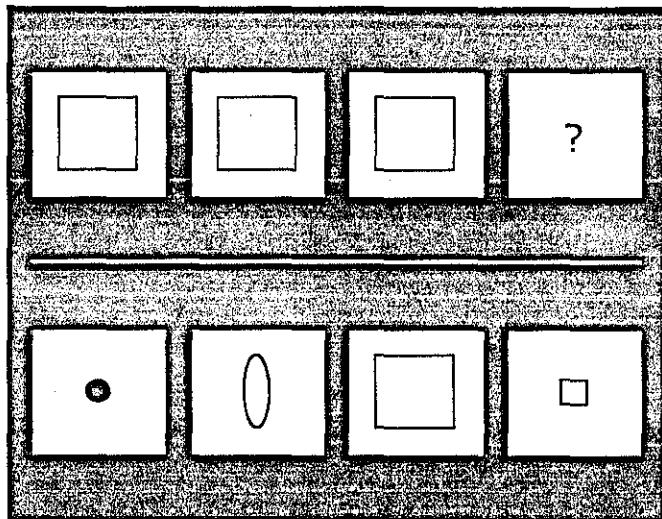
Subtest 'Seies' example 2
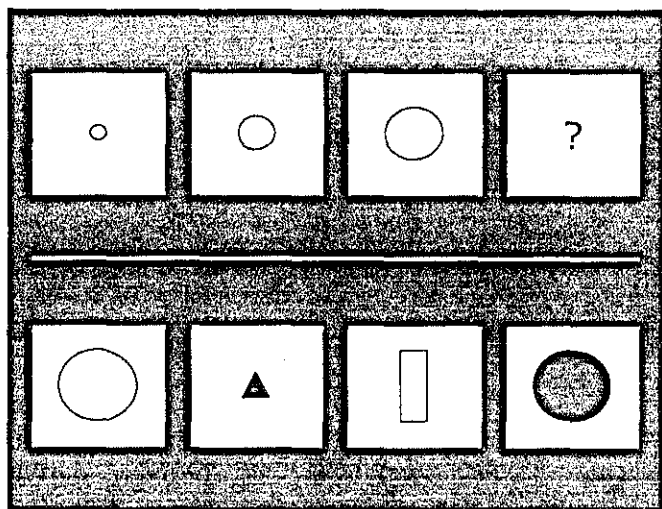
Subtest 'Series' example 3
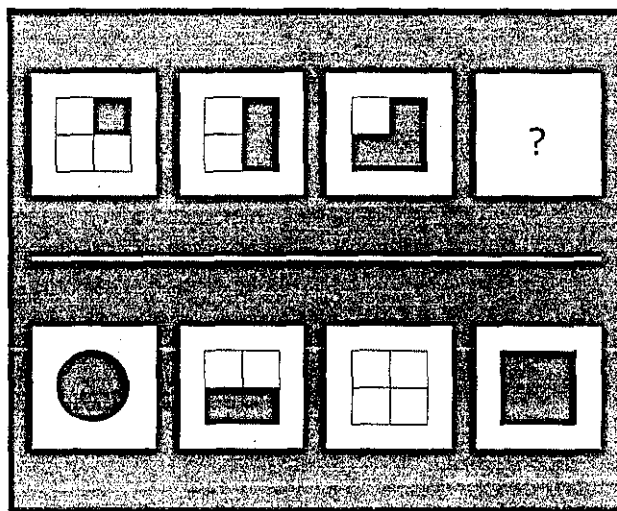
Subtest 'Odd One Out' example 1
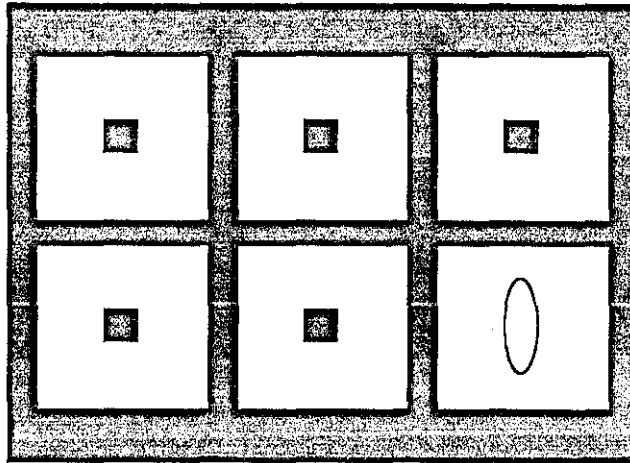
Subtest 'Odd One Out' example 2
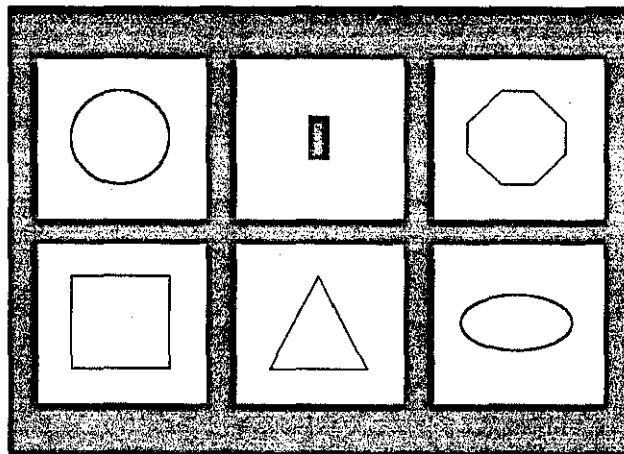
Subtest 'Odd One Out' example 3
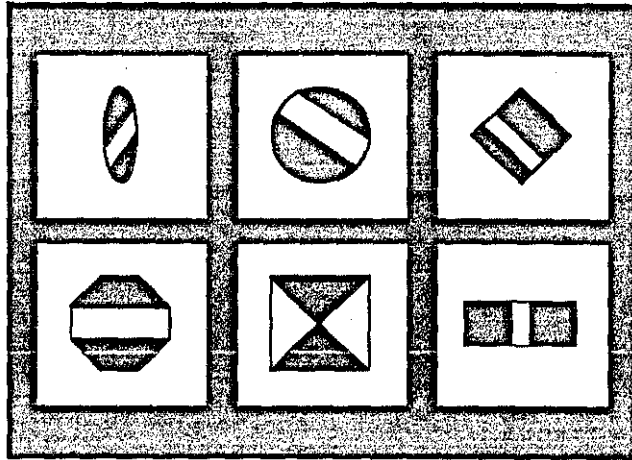
Subtest 'Jigsaw' example 1
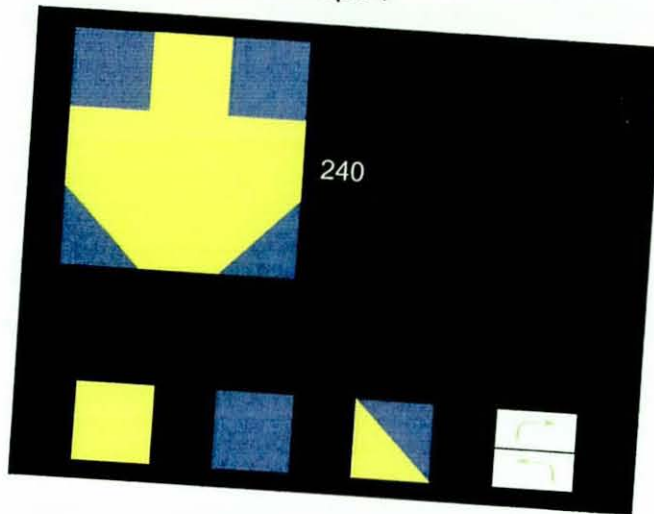
Subtest 'Jigsaw' example 2
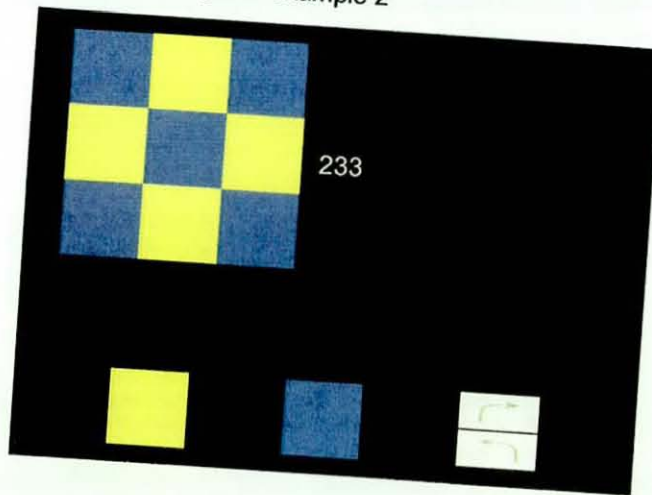
Subtest 'Jigsaw' example 3
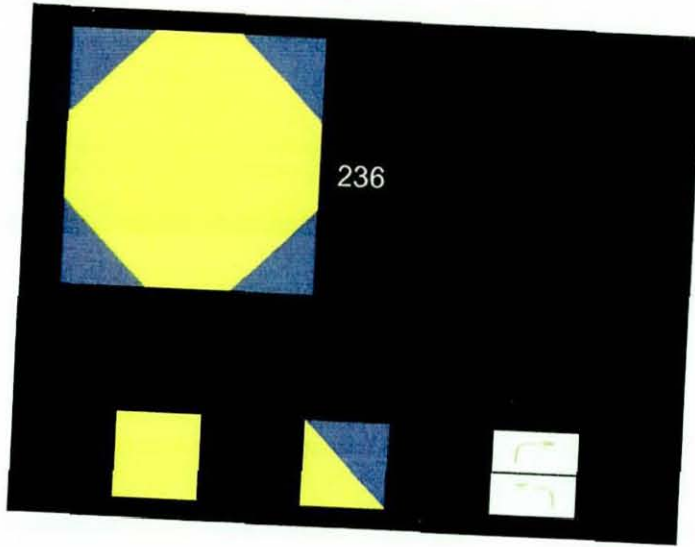
Appendix D: Instructions CCIID

Subtest 'Odd One Out'

For the pantomimed gestures the test administrator would first encircle the five pictures excluding the 'Odd One Out' and then point at the 'Odd One Out' to indicate that this object does not belong with the others. The participant would select the 'Odd One Out' by pointing to it with his finger (but without touching the computer screen in case of screens reacting to finger touch instead of computer pens). Only when the correct option was chosen, would the test administrator encourage the participant to touch the screen or hand the computer pen over to the participant. Then, the participant would be encouraged to point at the (correct) option again, using the pen. This was important because if the participant used the computer pen to point at an incorrect option, the test administrator would not have time to correct the mistake. For computer screen interfaces working by finger touch, the test administrator would put his hand in front of the screen to stop the participant from making a wrong choice and to correct the mistake. The pantomimed instructions remained the same for all three examples.

The verbal instructions for the first example were: *Here you see six shapes, five of them belong together, and one of them is different. Which one is different?* If the native language of the participant was not English, and the test administrator was not fluent in the participant's native language, the test administrator asked the coach to translate the verbal instructions. The verbal instructions for the second example would emphasise that all shapes are now different: *Now the shapes are all different but one is very different. Which one is it?* For the third example the test administrator would point out that although all shapes were very different, now they would share a common feature: *Now all shapes are very different, but they have one thing in common, but one shape does not. Which one is that? Which one is the 'Odd One Out'?* During

the examples, the coach was allowed to help to encourage the participant and was given the possibility to explain the instructions in the native language of the participant.

Depending on the response of the participant the test administrator would react accordingly:

If the participant picked the right shape, the test administrator would say: *Well done!* and would move on the next example, or, after the last example, to the first test version.

If the participant would choose the incorrect shape, the test administrator would first encourage the participant to choose again, and if the second choice was incorrect again, point out the correct shape: *This one is different!*

If the participant did not react despite encouragement, the test administrator would point out the 'Odd One Out' shape to the participant. Again, the test administrator should encircle the five shapes excluding the 'Odd One Out' and then point to the 'Odd One Out' and say: *This one is different.* (if necessary translated by the coach).

If the participant pointed to two or more shapes the test administrator would say: *Only one is most different. Which one is it?* (if necessary translated by the coach).

If necessary, basic test instructions were used after the examples if necessary to remind the participant what to do (*'Which one is most different?'*). The test administrator would not provide corrections after the completion of the examples.

Subtest 'Series'

Basic pantomimed gestures were used to explain the task of the subtest supported by examples. The test administrator would first point at the upper row of shapes and then at the question mark at the end of the row. Next, the test administrator would point over the options and go back to the question mark and, then, give the participant time to choose an option. The participant would select the option by pointing to it with his finger. The pantomimed instructions remained the same for all three examples.

The verbal instructions for the first example were: *In the row up here, you see three shapes and a question mark* (point). *Which one of these shapes* (point to the options) *belongs in the place of the question mark?*

The verbal instructions for examples two and three were: *Now these shapes are changing* (point to the upper row). (Point to first shape) *This changes into that* (point to second shape) *and that changes further into that* (point to third shape) *and that changes further again* (point to question mark). *How would it look like out of those?* (point to options).

If the answer was correct, the test administrator would say: *Well done!* The test administrator would move on to the next example or, after the last example, to the first test version.

If the answer was incorrect, the test administrator would first encourage the participant to choose again, and, if the second choice was incorrect again, point out the correct shape: *Look, it is this one.* (Points to the first shape and explain again:) *This changes into that* (points to second shape) *and that changes further into that* (points to third shape) *and that changes further into this one* (points to the correct option).

If the participant did not react despite encouragement, the test administrator would point out the correct option to the participant.

Test instructions were used after the examples, if deemed necessary. However, instructions would only be given to remind the participant what to do. The test administrator would not correct the participant after the completion of the examples.

Subtest 'Jigsaw'

While completing the first example the test administrator would explain the different features of the subtest when appropriate. First, the test administrator would point at the geometric picture on the left and say: *The task here is to copy this picture into that frame* (pointing at empty frame) *using these pieces* (pointing at pieces). *You move them like this* (the examiner starts filling the frame, makes a mistake on purpose in upper row but continues to fill in pieces into the first two rows from the top).

After 2nd row: *You can only change the position of the last piece* (test assistant demonstrates that only the piece that had just been moved into the frame can be moved again, not the others). *You can see, I made a mistake here. I can't change these pieces anymore, so I just put new ones over it* (test assistant corrects mistake). *If you want to change a piece, just put a new piece over it*

Before starting on the third row the test assistant points out the patterned pieces in the picture on the left and points at the patterned piece on the bottom of the screen: *You also need patterned pieces for some pictures* (test assistant points at picture) *like this one which you can turn* (points at piece at the bottom of screen). *When you tick that arrow* (test assistant points out arrow) *you can turn the patterned piece into the right position* (test assistant moves pieces into the right position and shows rotation)

When the frame is full, this arrow appears (points at arrow between the two pictures). *It does not mean that the picture is the same as this one (point) it just means the frame is full. Just check if the two pictures are the same if this is so, tick this arrow and the next picture will appear.*

After the first example the test assistant would say: *Now you try, please.* Following the first example, the test assistant would hand the pen over to the participant to continue with example two and three. Instructions

were repeated and supported by the sports coach when necessary. Corrections were made if needed.

After the examples, instructions were given to remind the participant of the task (*Please copy this picture into that frame*) and to reiterate how to correct mistakes if the participant tried to change pieces (*Use a new piece*). No corrections were given after the example items.

Appendix E: Subtests CCIID

Subtest 'Series' version 1 item 1
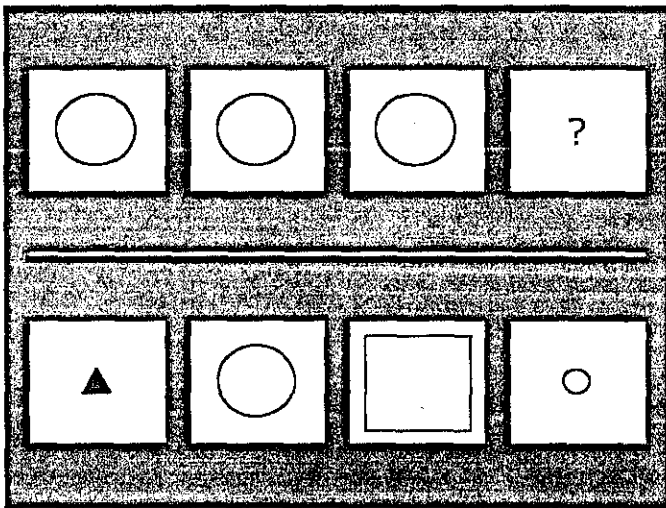
Subtest 'Series' version 1 item 2
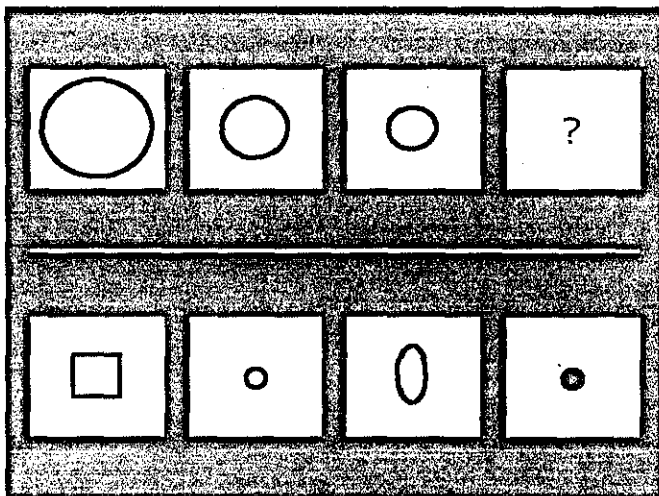
Subtest 'Series' version 1, item 3
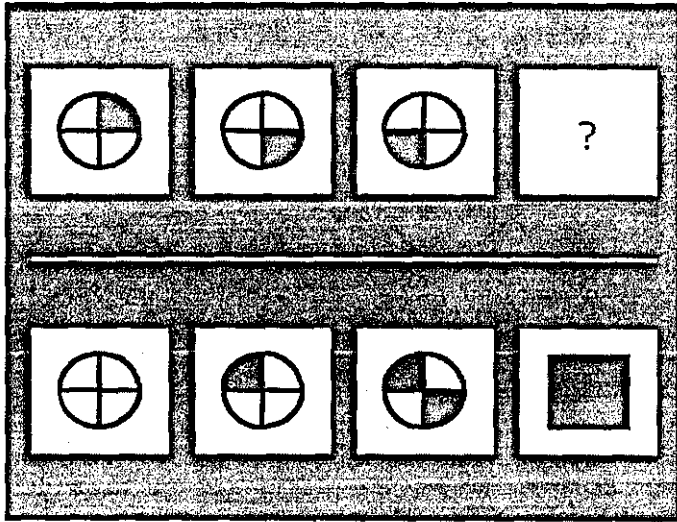
Subtest 'Series' version 1, item 4
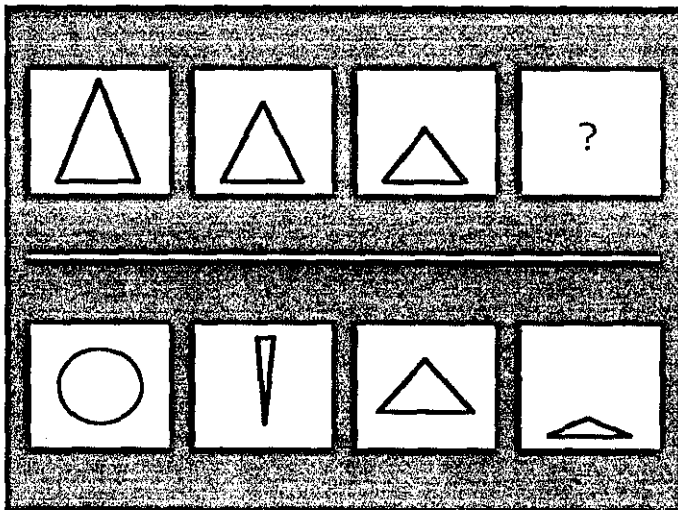
Subtest 'Series' version 1 item 5
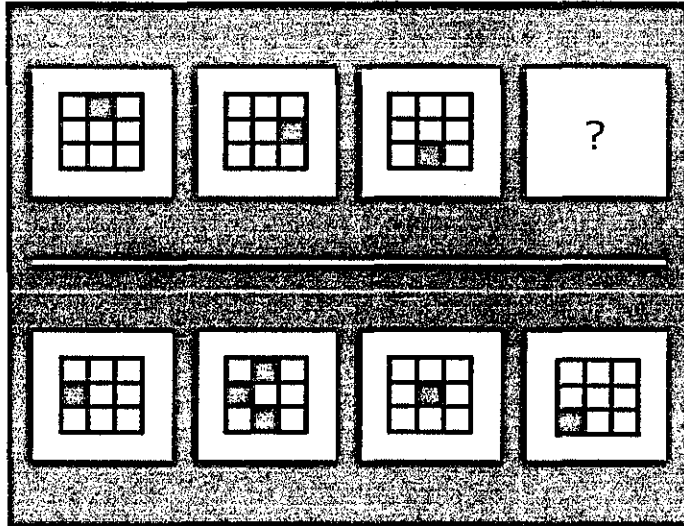
Subtest 'Series' version 1 item 6
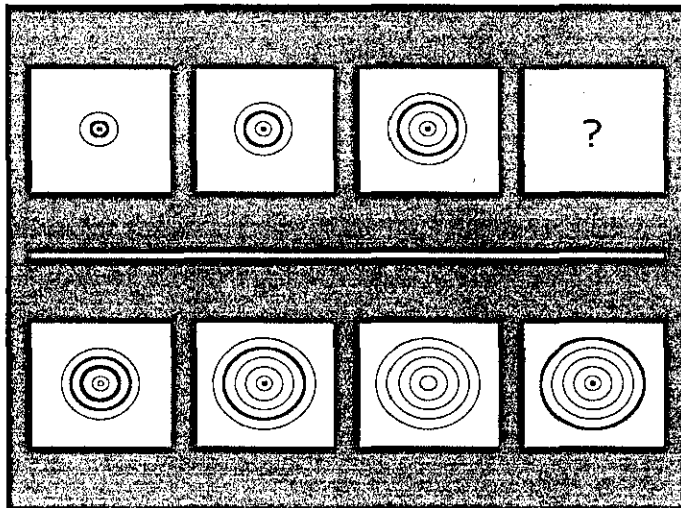
Subtest 'Series' version 1 item 7
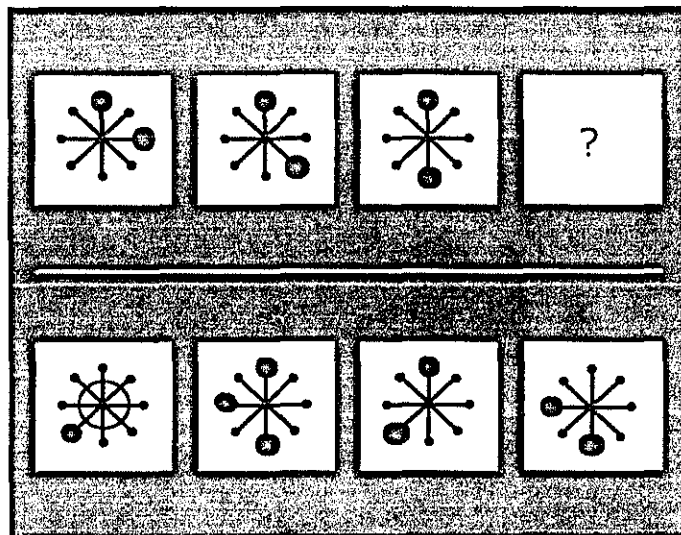
Subtest 'Series' version 1 item 8

247
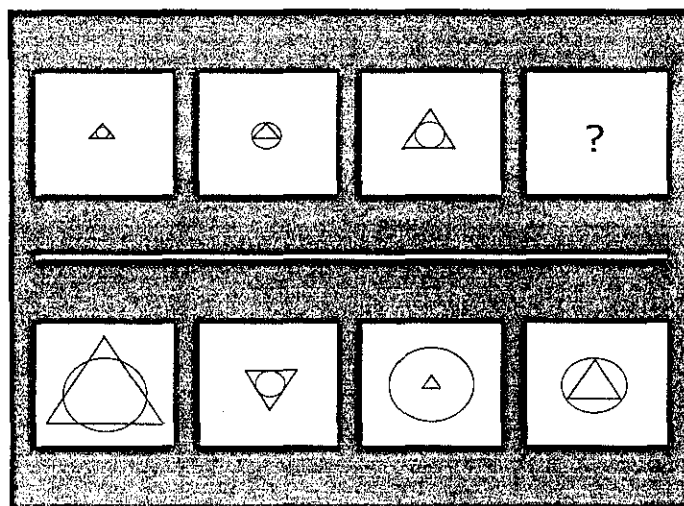
Subtest 'Series' version 1 item 9
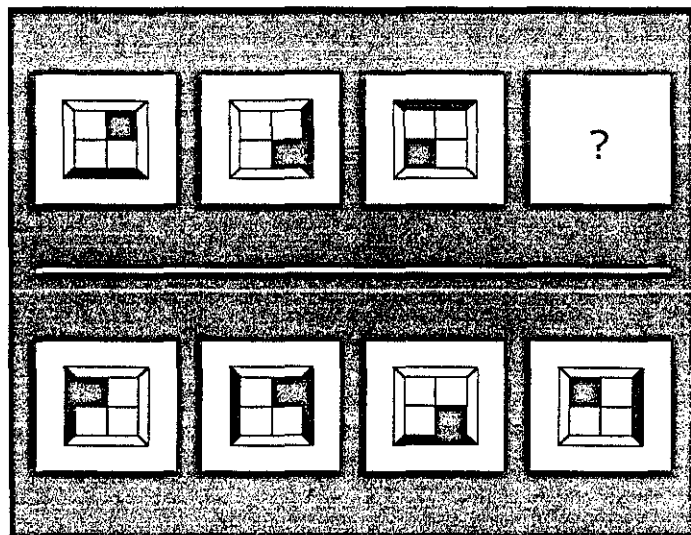
Subtest 'Series' version 1 item 10
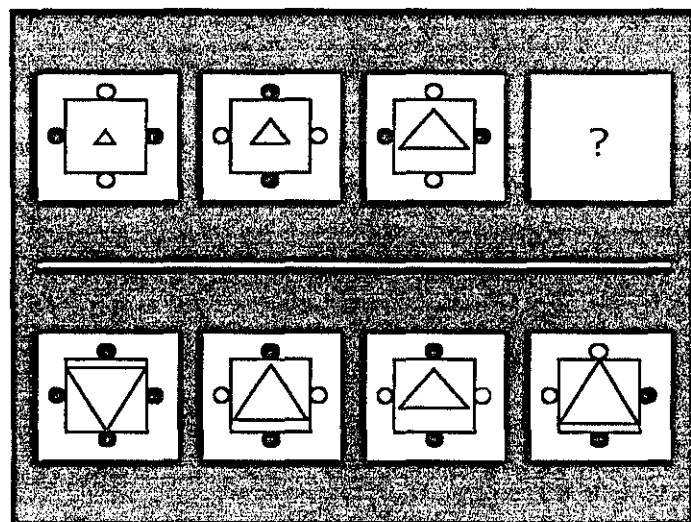
Subtest 'Series' version 1 item 11
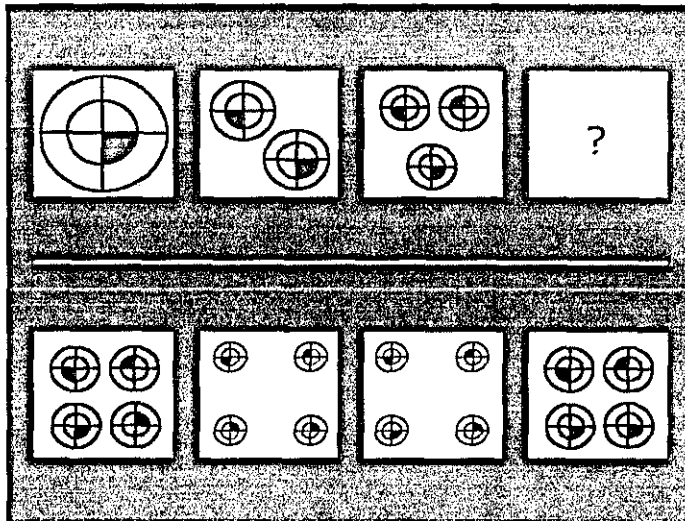
Subtest 'Series' version 1 item 12
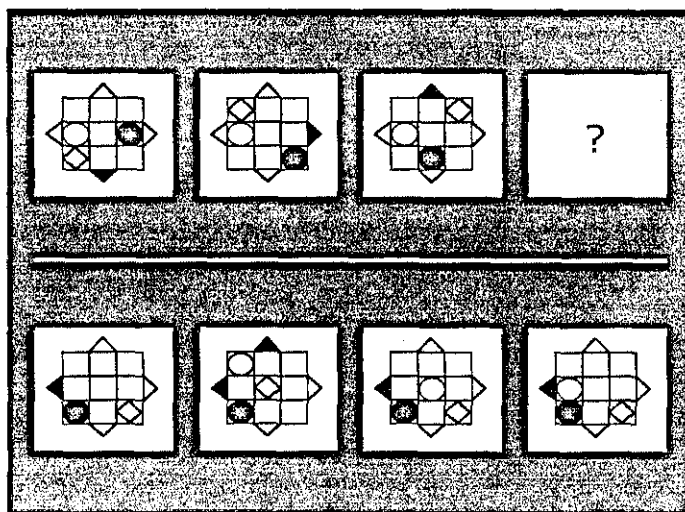
Subtest 'Series' version 1 item 13
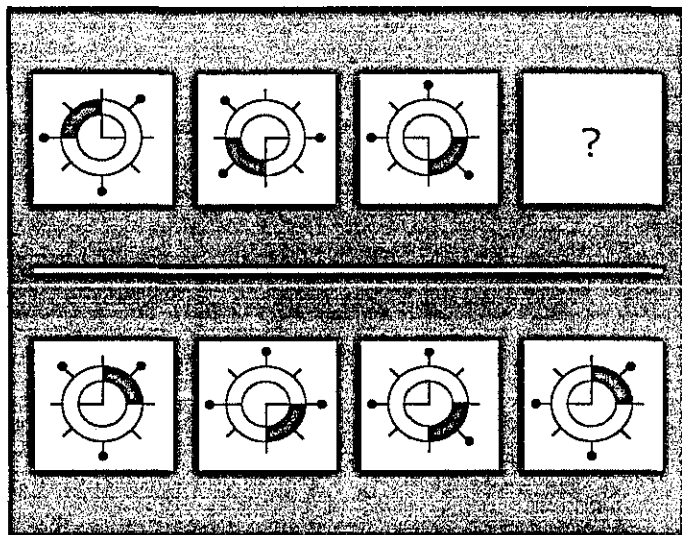
Subtest 'Series' version 2 item 1
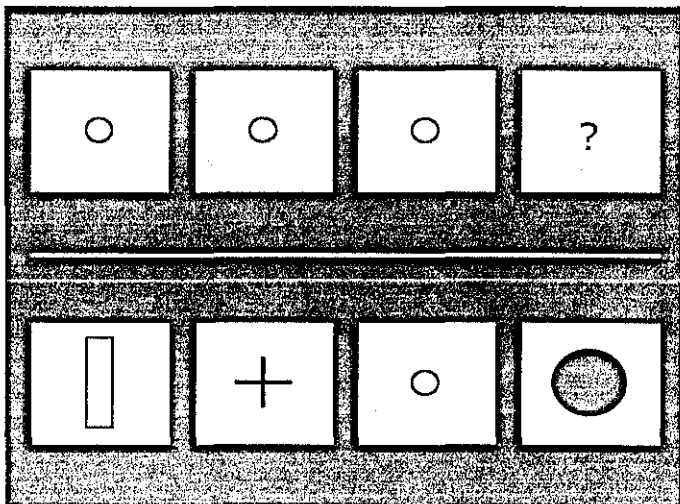
Subtest 'Series' version 2 item 2

251
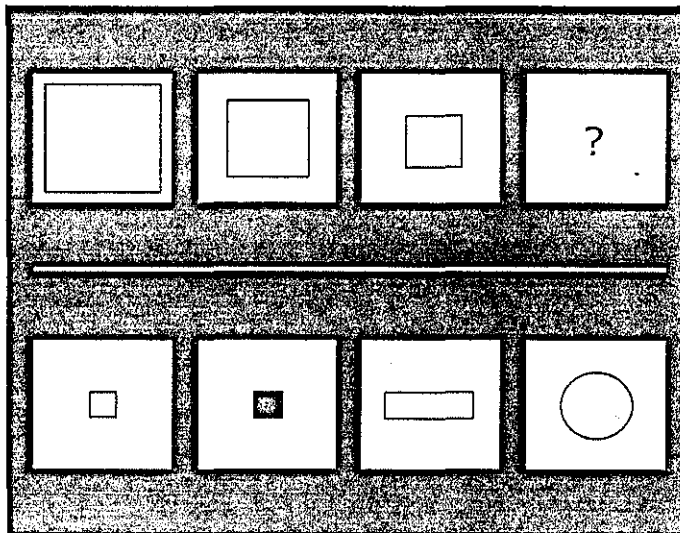
Subtest 'Series' version 2 item 3
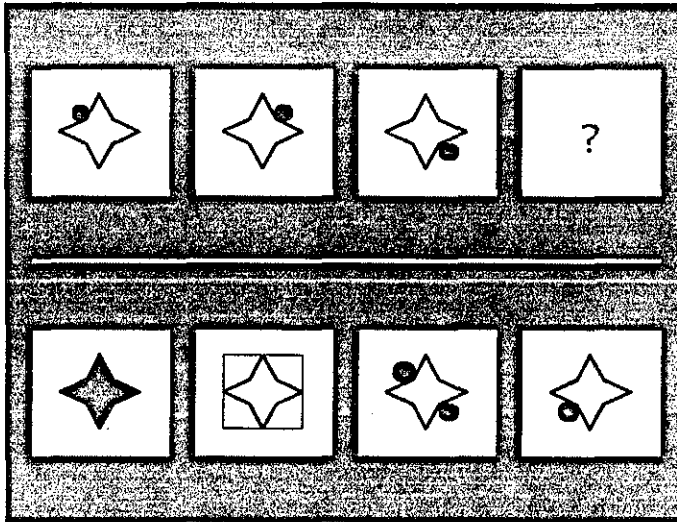
Subtest 'Series' version 2 item 4

252
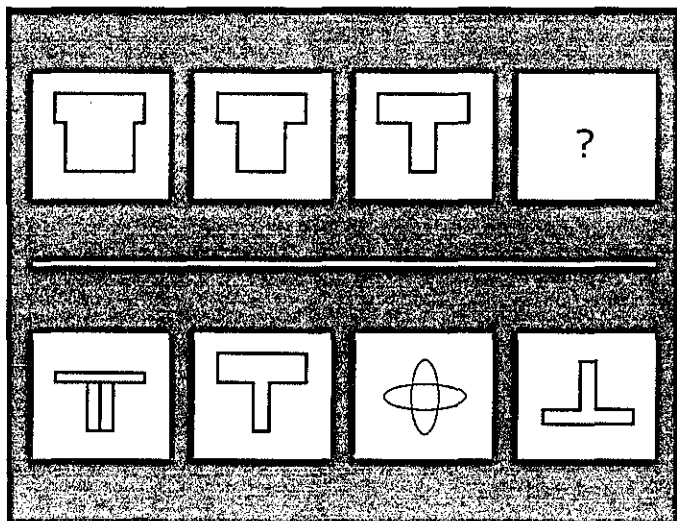
Subtest 'Series' version 2 item 5
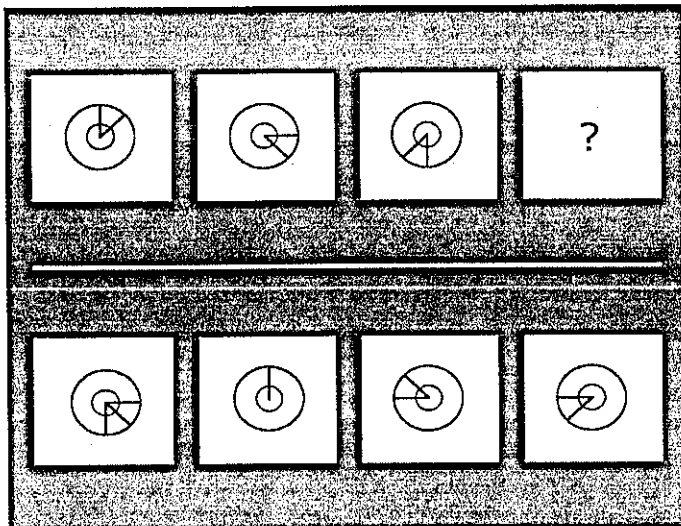
Subtest 'Series' version 2 item 6

253
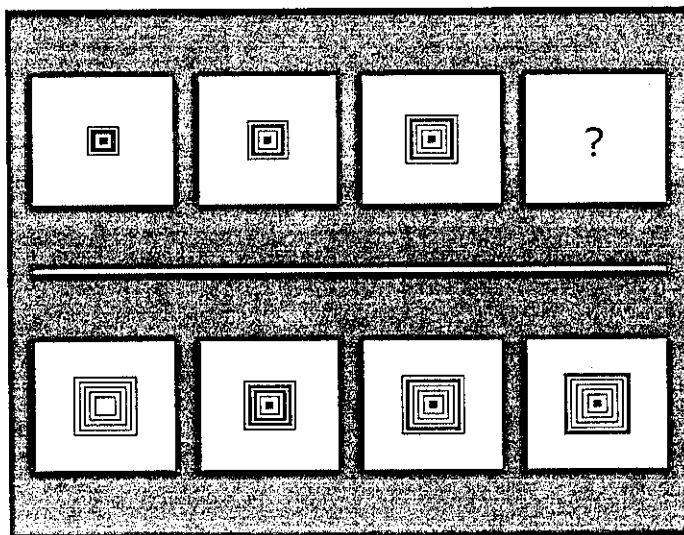
Subtest 'Series' version 2 item 7
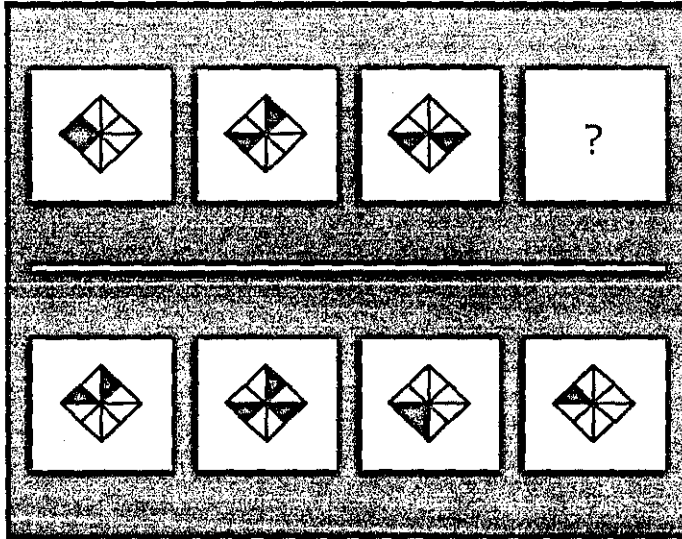
Subtest 'Series' version 2 item 8

254
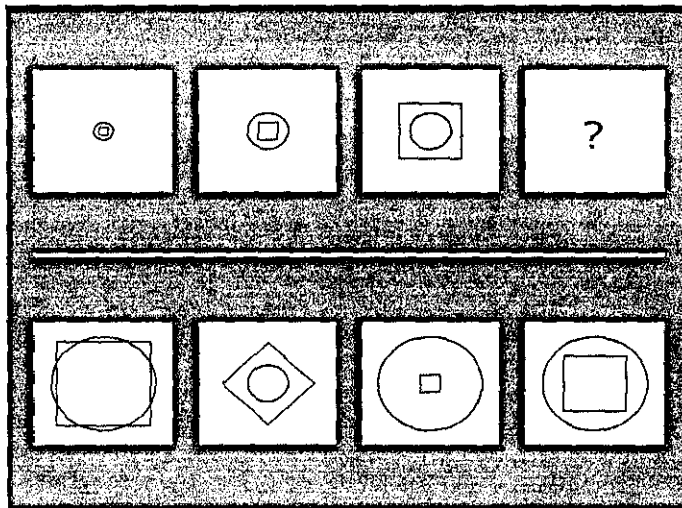
Subtest 'Series' version 2 item 9
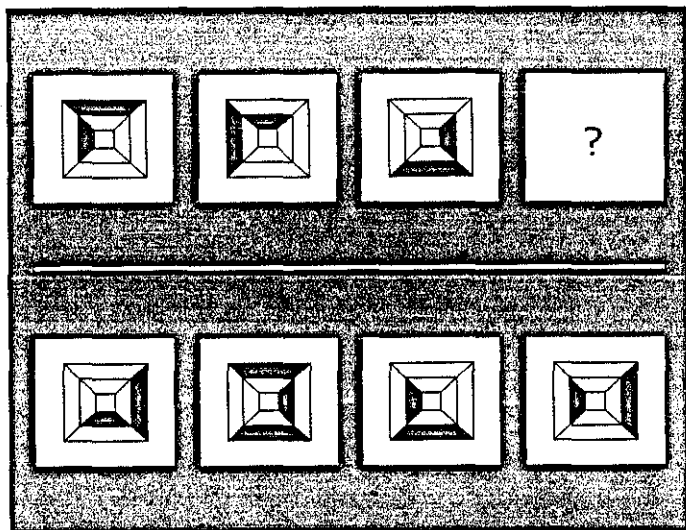
Subtest 'Series' version 2 item 10

255
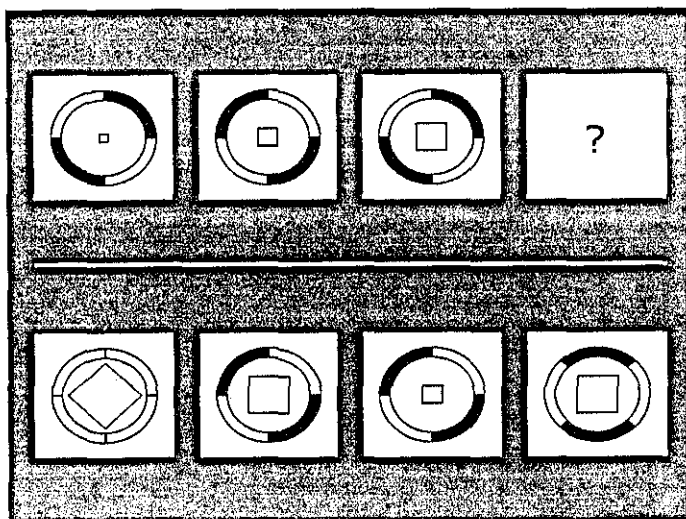
Subtest 'Series' version 2 item 11
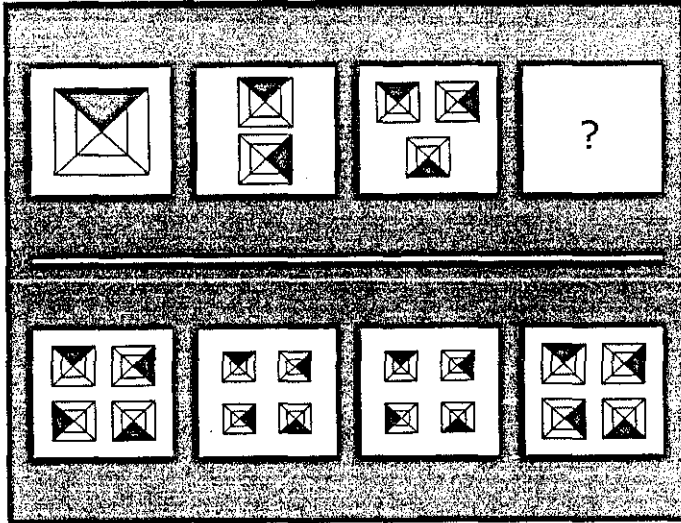
Subtest 'Series' version 2 item 12

256
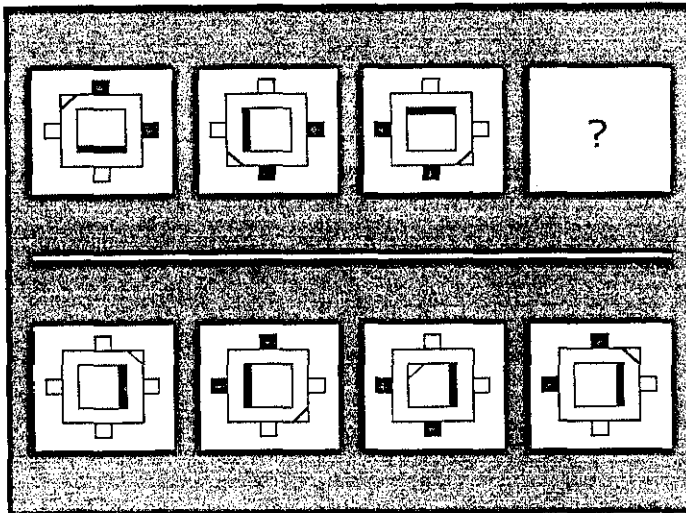
Subtest 'Series' version 2 item 13
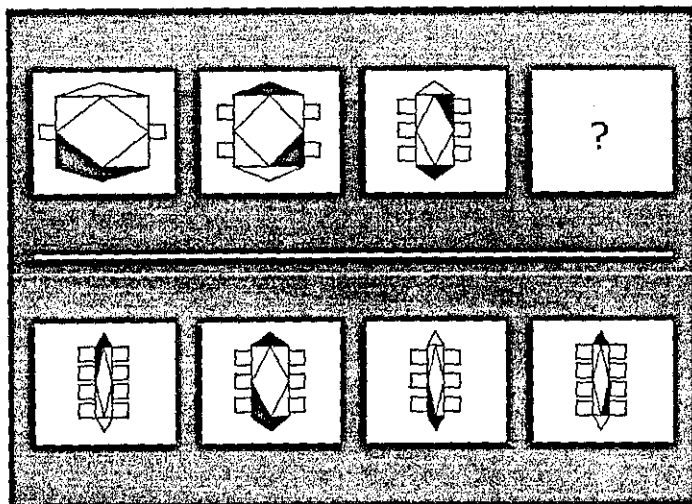
Subtest 'Series' version 3 item 1
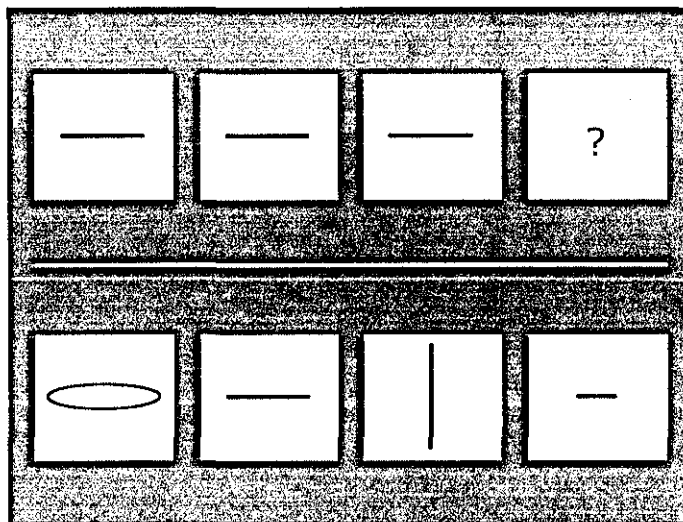
Subtest 'Series' version 3 item 2

258
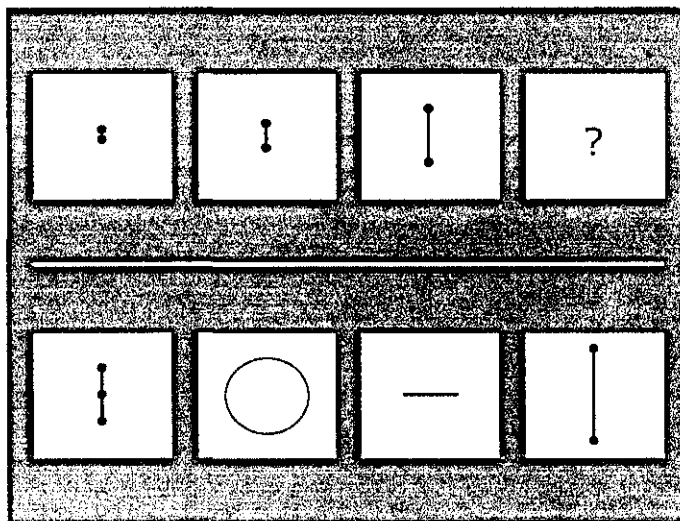
Subtest 'Series' version 3 item 3
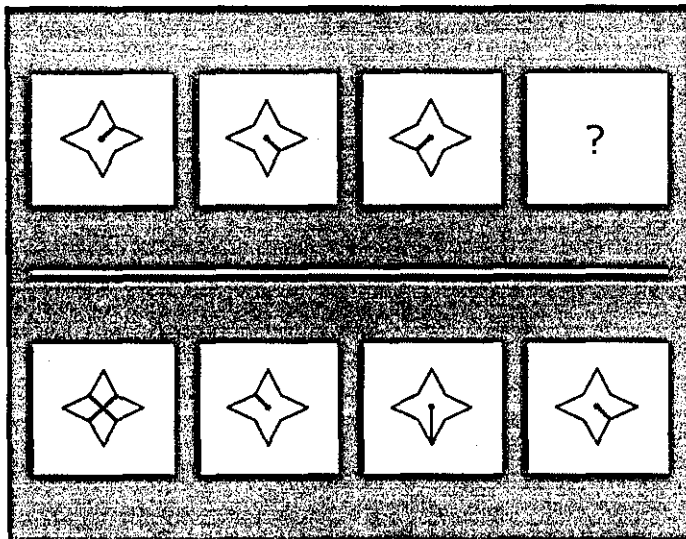
Subtest 'Series' version 3 item 4

259
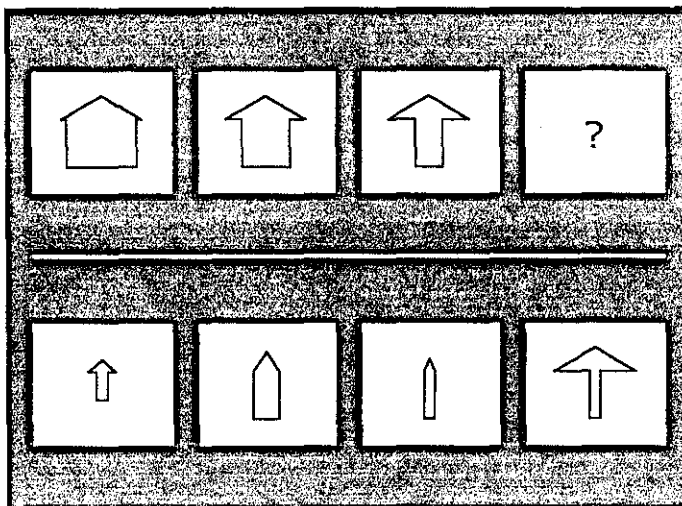
Subtest 'Series' version 3 item 5
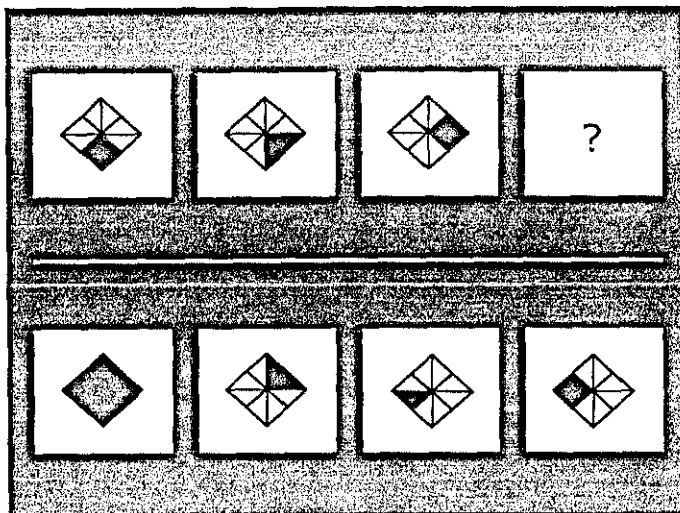
Subtest 'Series' version 3 item 6

260
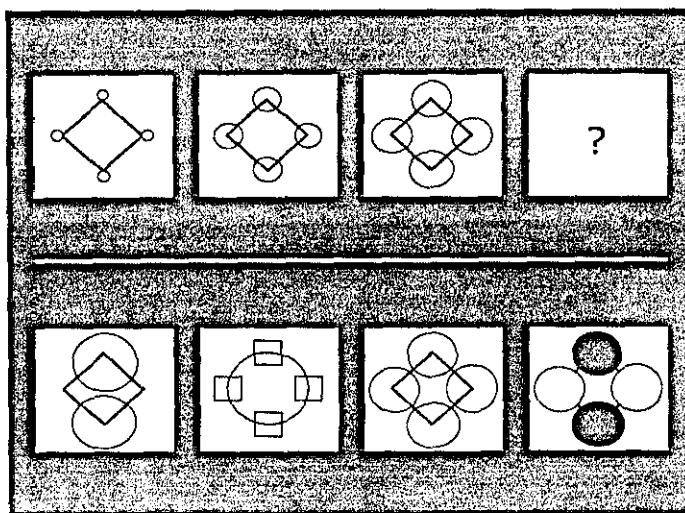
Subtest 'Series' version 3 item 7
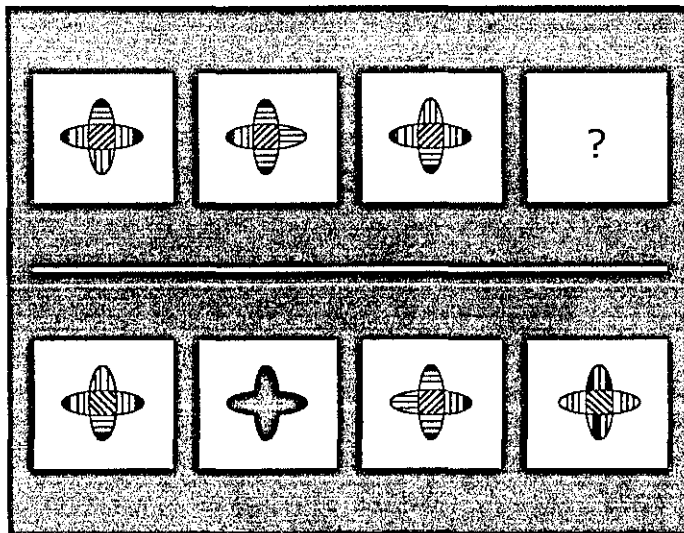
Subtest 'Series' version 3 item 8

261
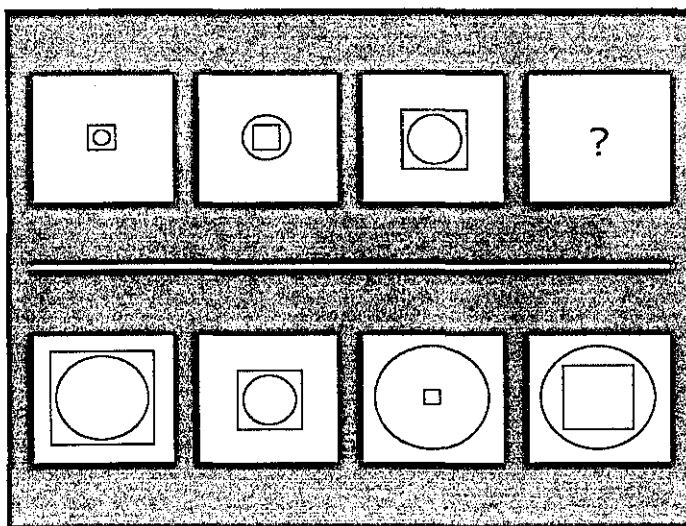
Subtest 'Series' version 3 item 9
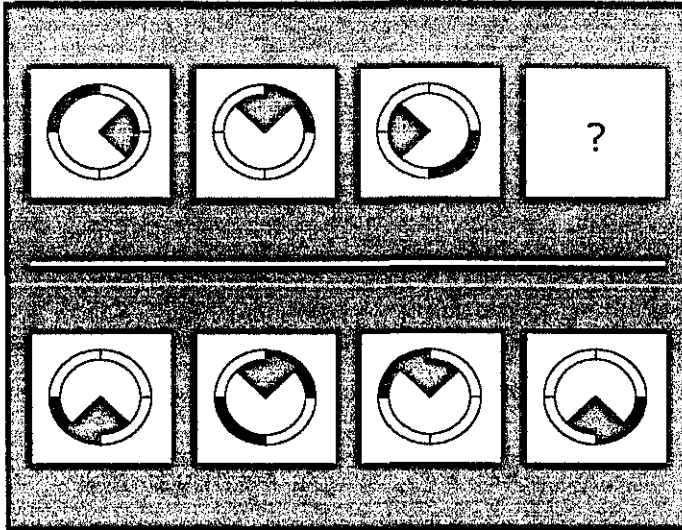
Subtest 'Series' version 3 item 10
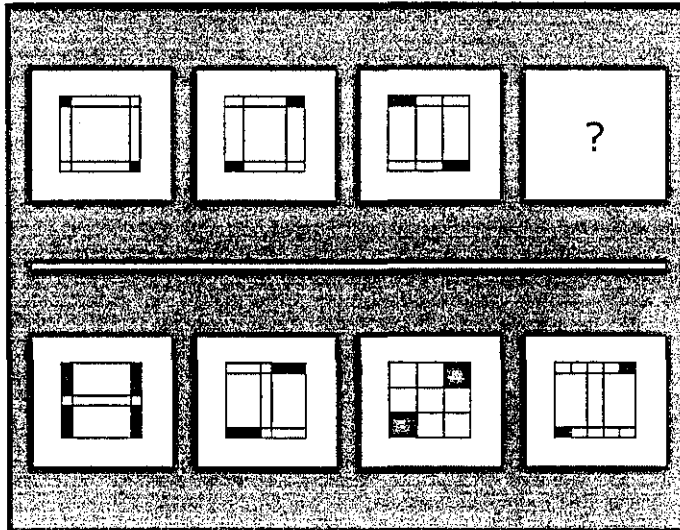
Subtest 'Series' version 3 item 11
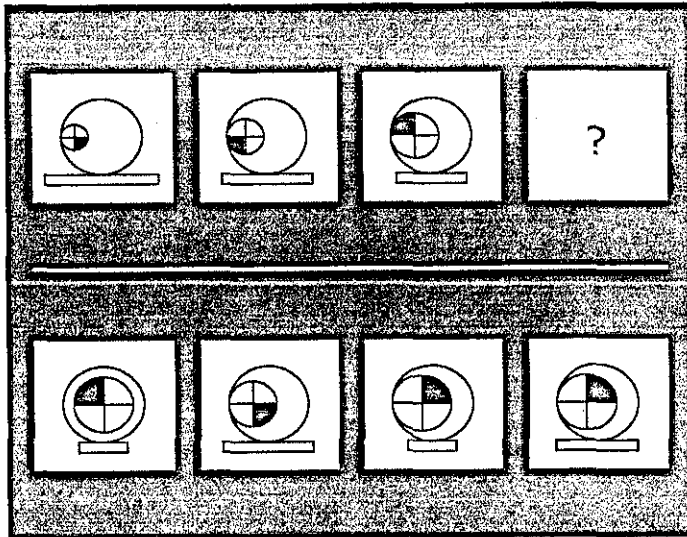
Subtest 'Series' version 3 item 12
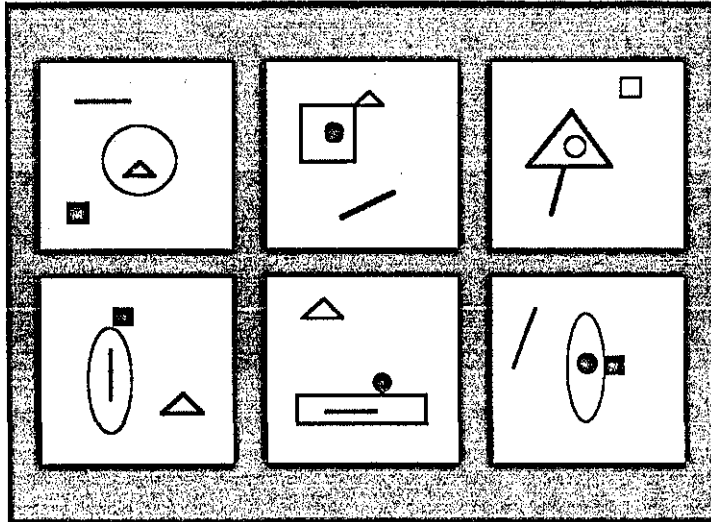

263
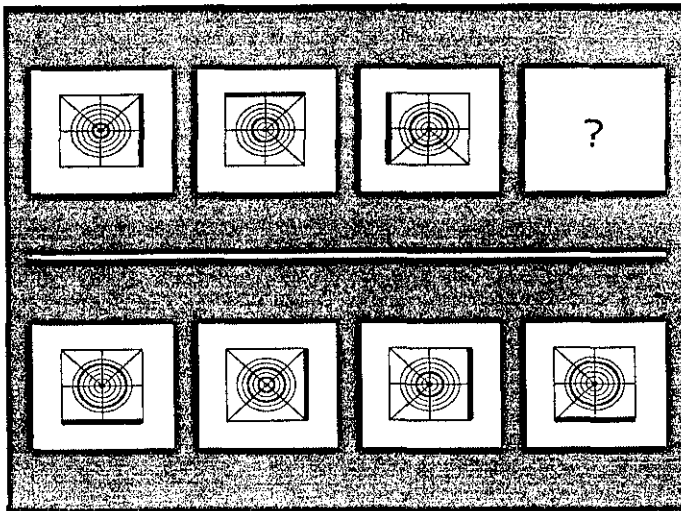
Subtest 'Series' version 3 item 13
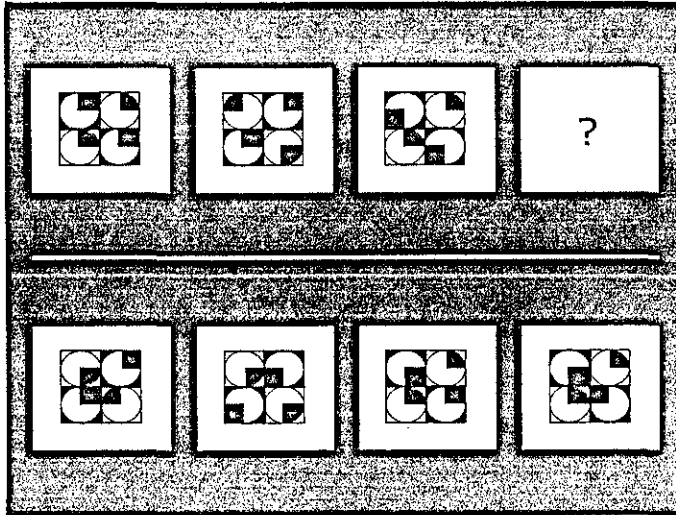
Subtest 'Odd One Out' version 1, item 1
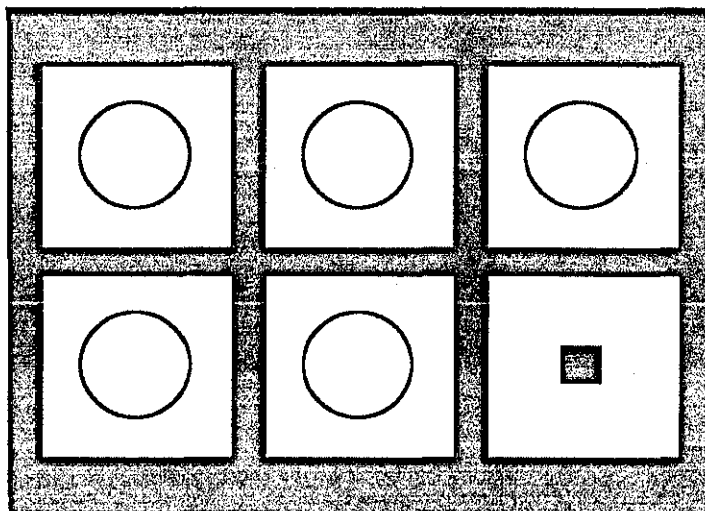
Subtest 'Odd One Out' version 1, item 2
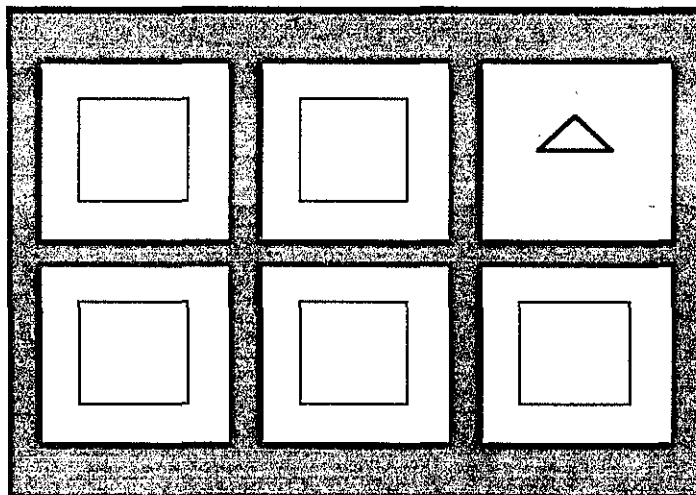
Subtest 'Odd One Out' version 1, item 3
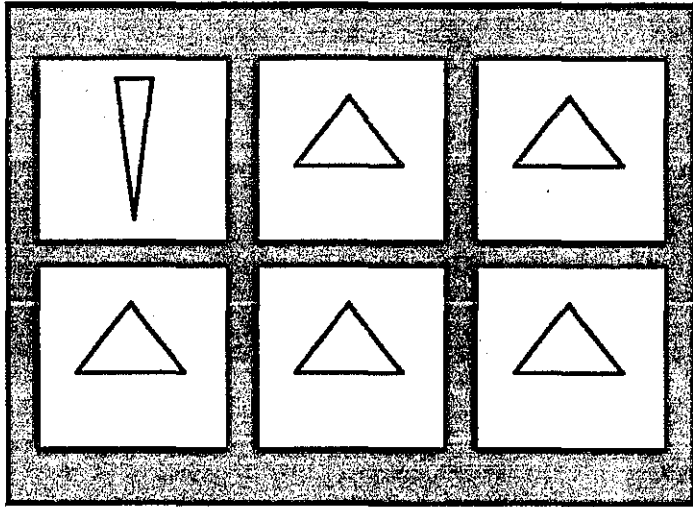
Subtest 'Odd One Out' version 1, item 4
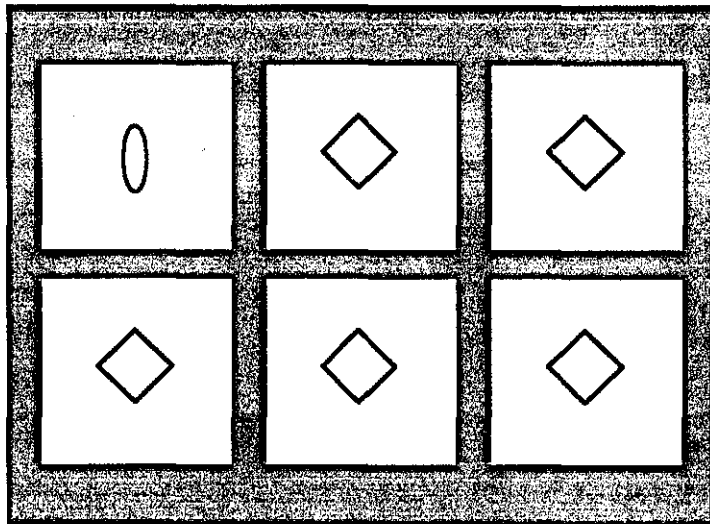
Subtest 'Odd One Out' version 1, item 5
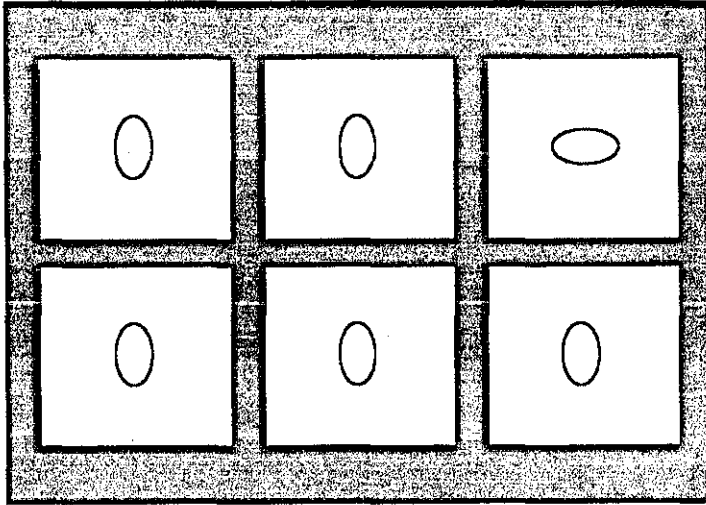
Subtest 'Odd One Out' version 1, item 6
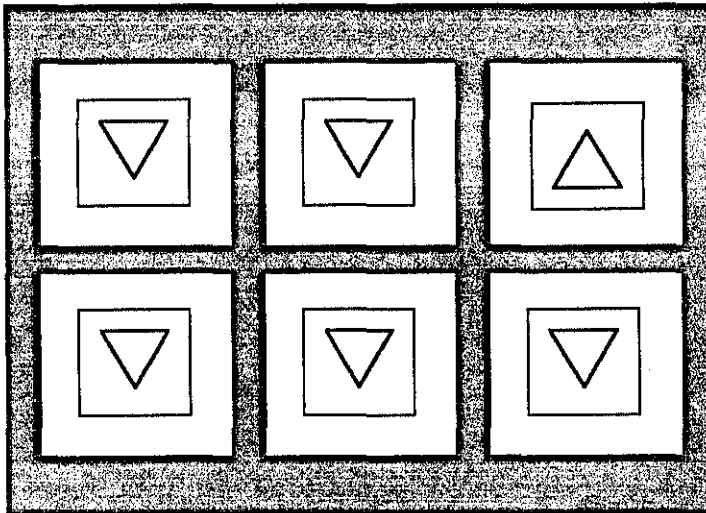
Subtest 'Odd One Out' version 1, item 7
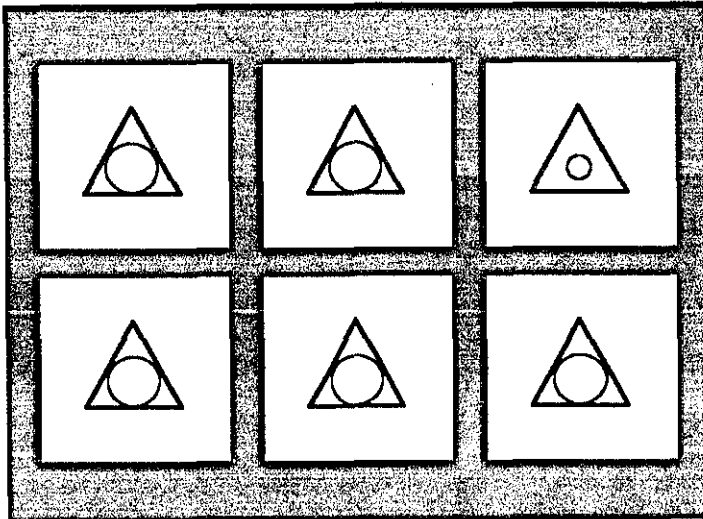
Subtest 'Odd One Out' version 1, item 8
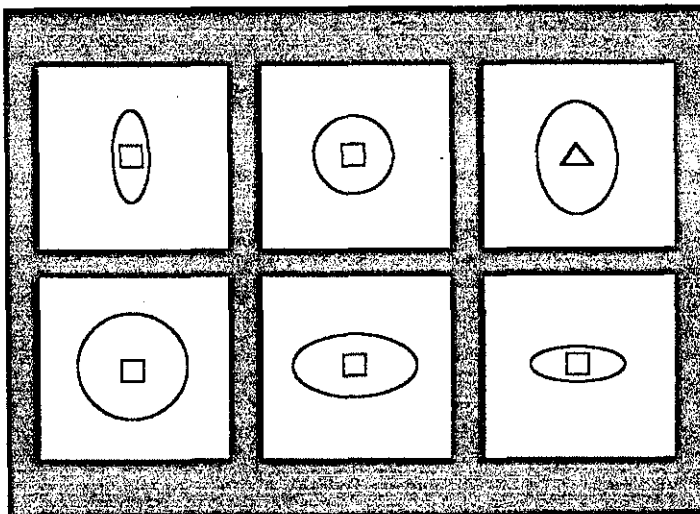
Subtest 'Odd One Out' version 1, item 9
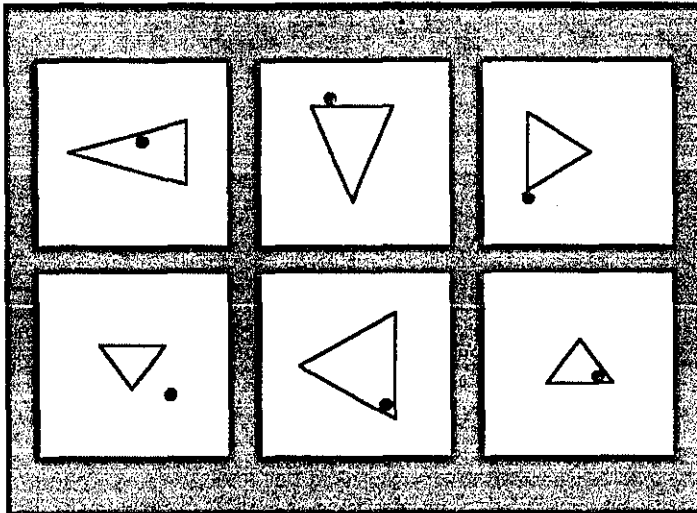
Subtest 'Odd One Out' version 1, item 10
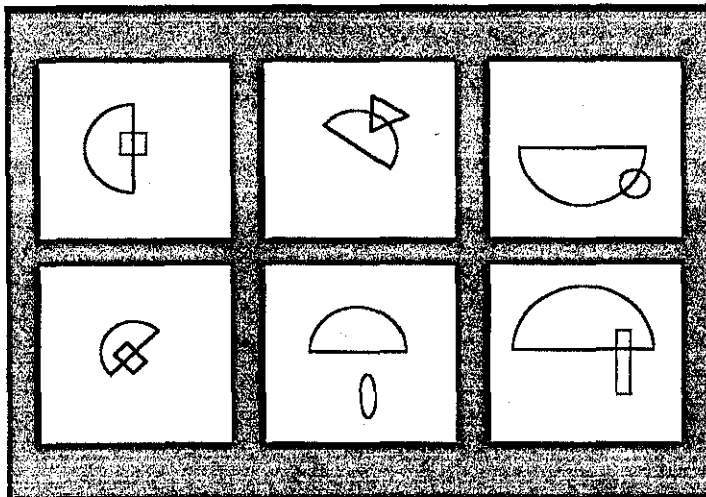
Subtest 'Odd One Out' version 1, item 11
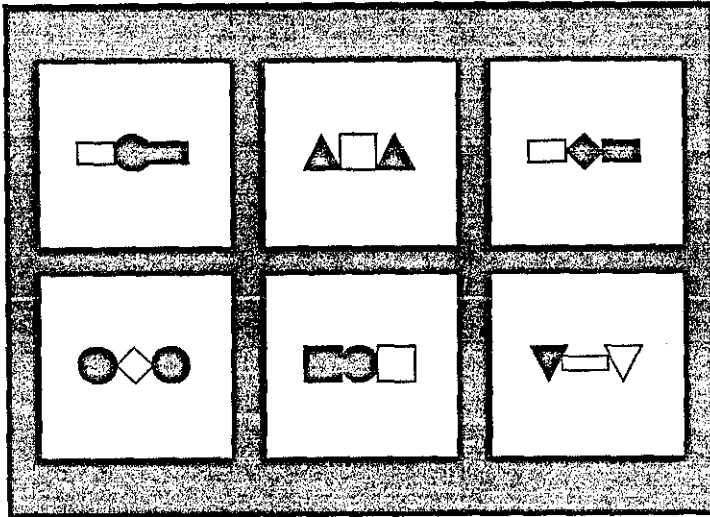
Subtest 'Odd One Out' version 1, item 12
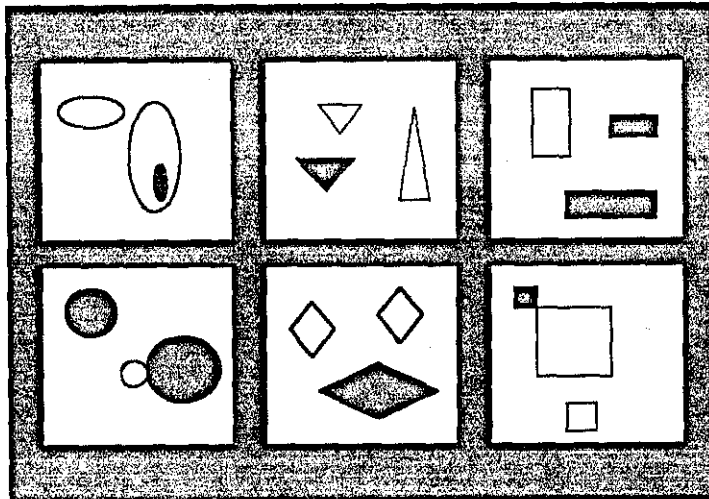
Subtest 'Odd One Out' version 1, item 13
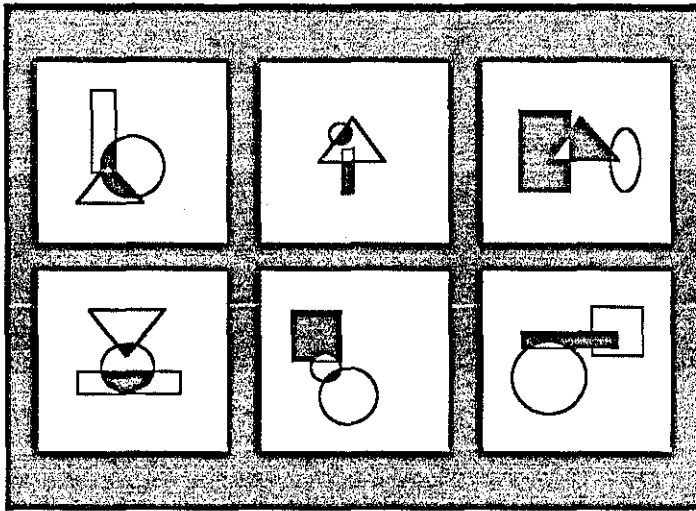

Subtest 'Odd One Out' version 1, item 14
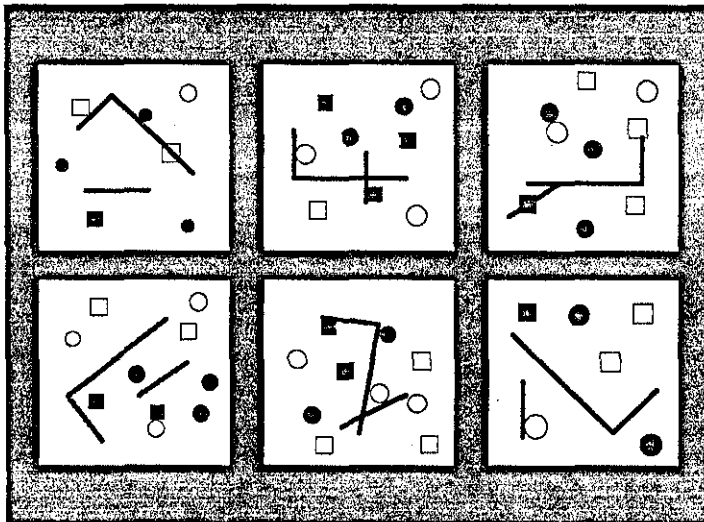
Subtest 'Odd One Out' version 2, item 1
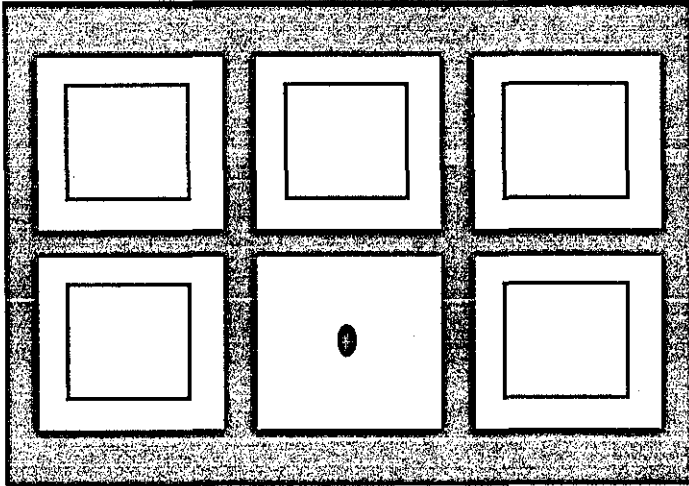
Subtest 'Odd One Out' version 2, item 2
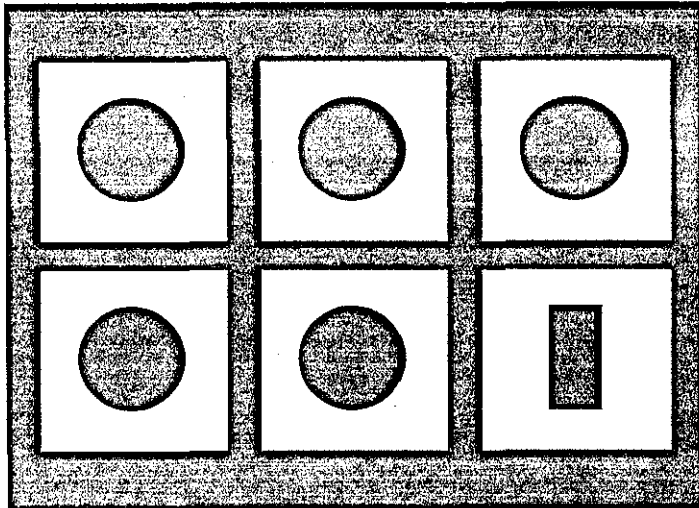
Subtest 'Odd One Out' version 2, item 3
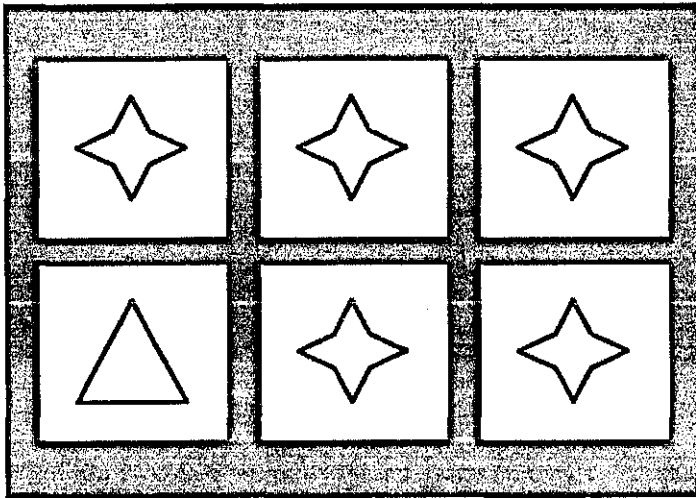
Subtest 'Odd One Out' version 2, item 4
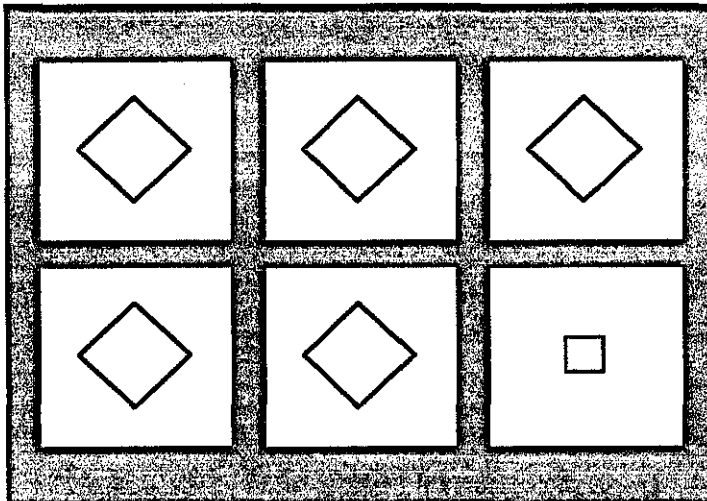
Subtest 'Odd One Out' version 2, item 5
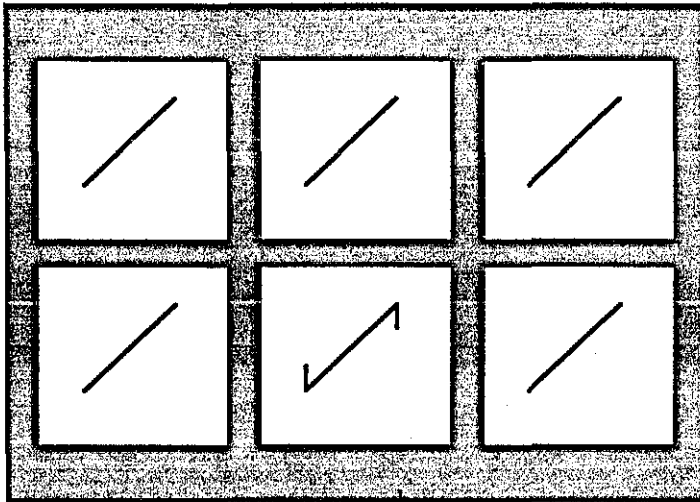

Subtest 'Odd One Out' version 2, item 6
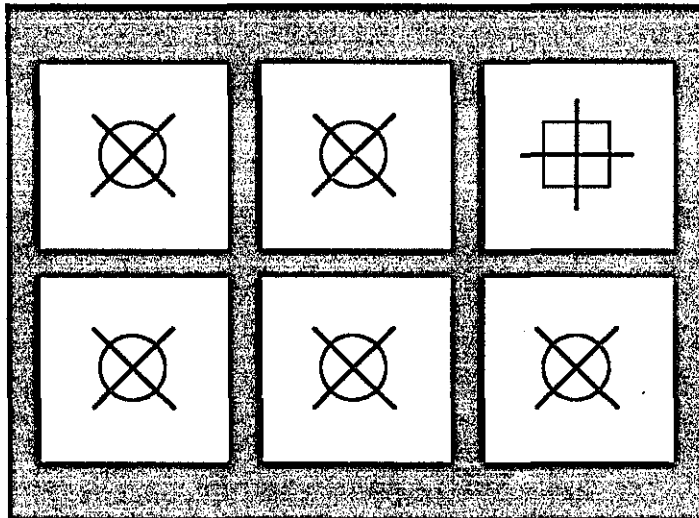
Subtest 'Odd One Out' version 2, item 7
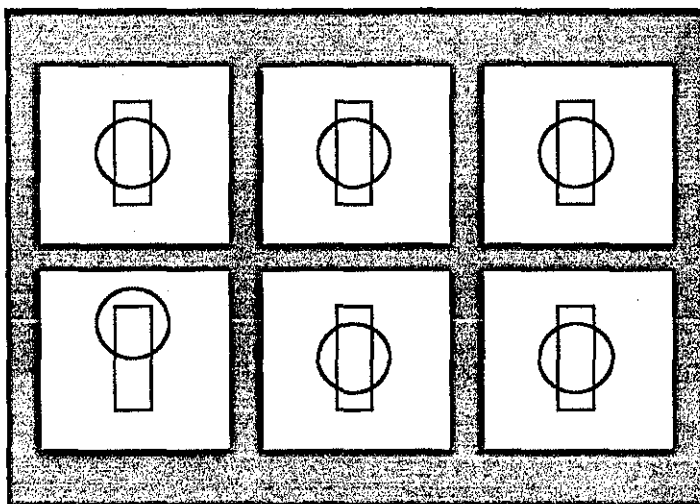
Subtest 'Odd One Out' version 2, item 8
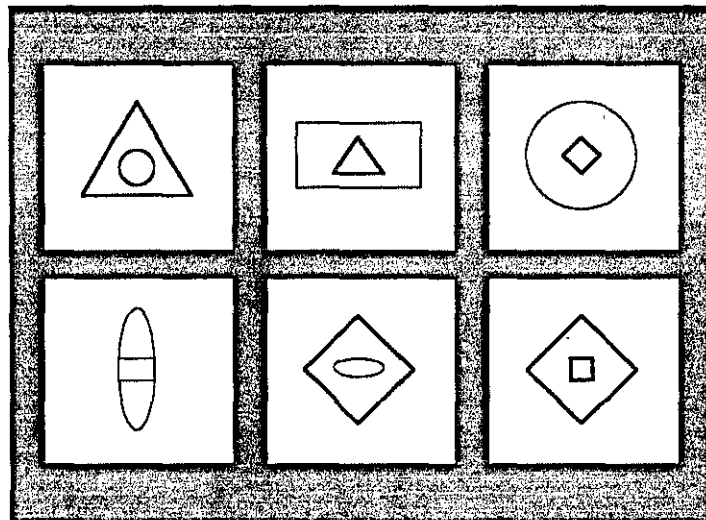
Subtest 'Odd One Out' version 2, item 9
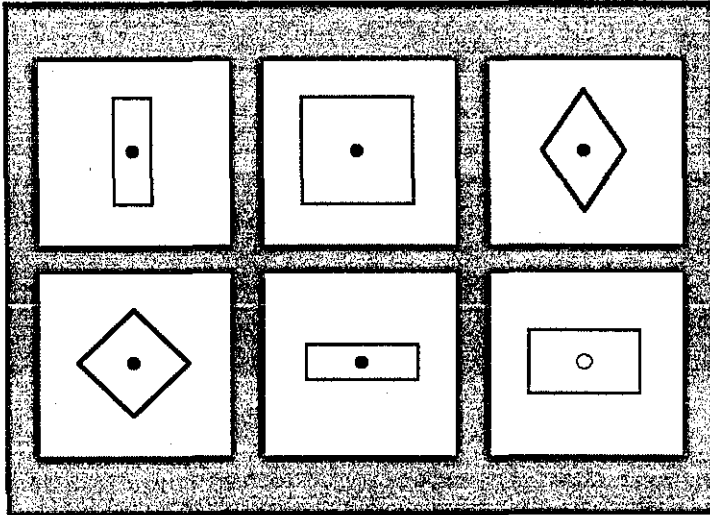
Subtest 'Odd One Out' version 2, item 10
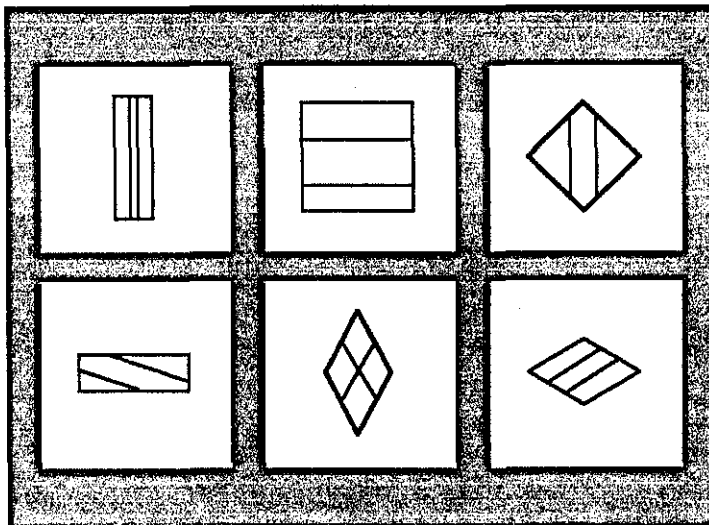
Subtest 'Odd One Out' version 2, item 11
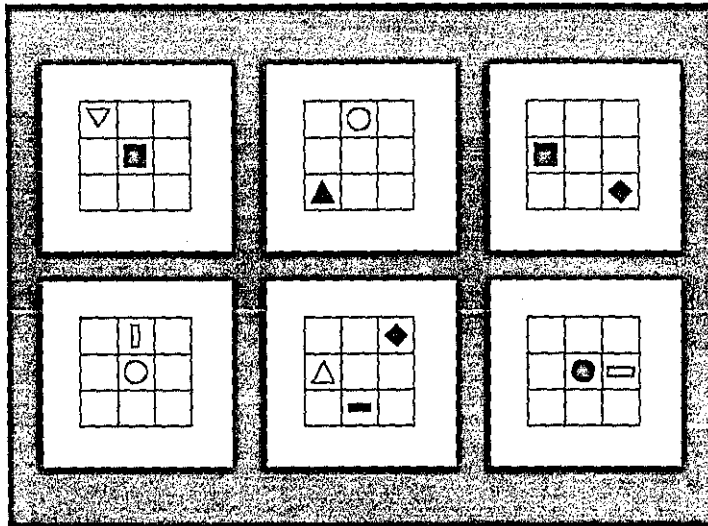
Subtest 'Odd One Out' version 2, item 12
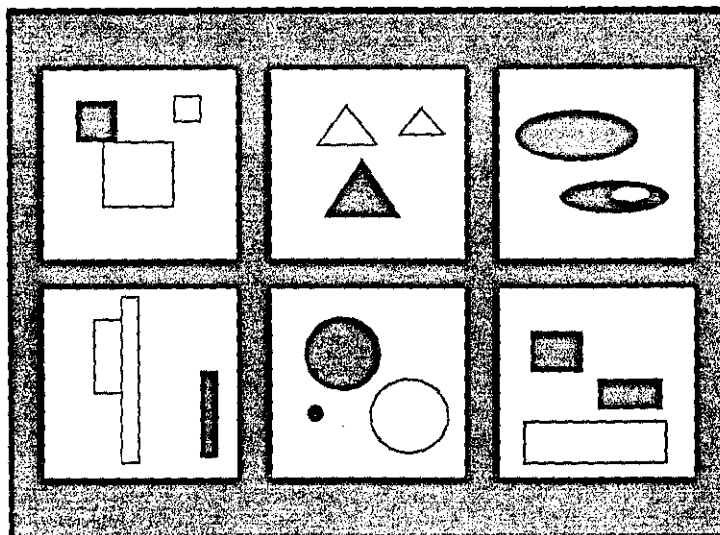
Subtest 'Odd One Out' version 2, item 13
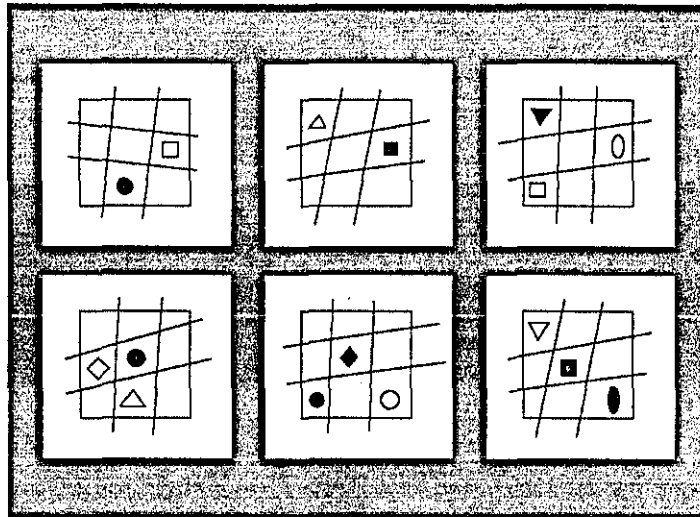
Subtest 'Odd One Out' version 2, item 14
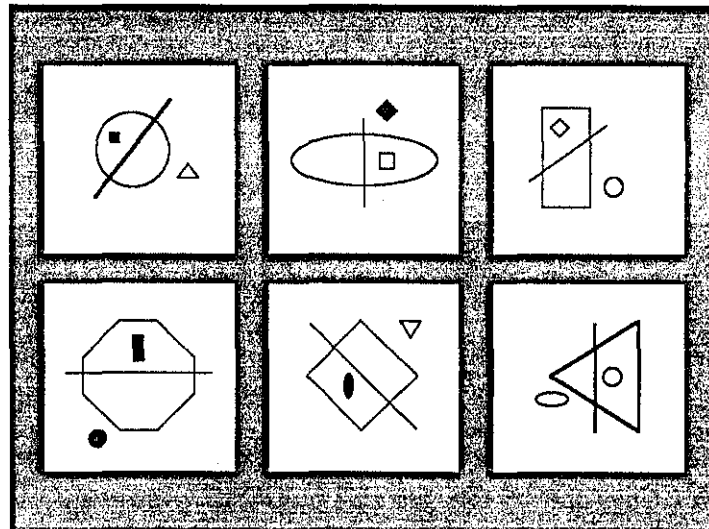
Subtest 'Odd One Out' version 3, item 1
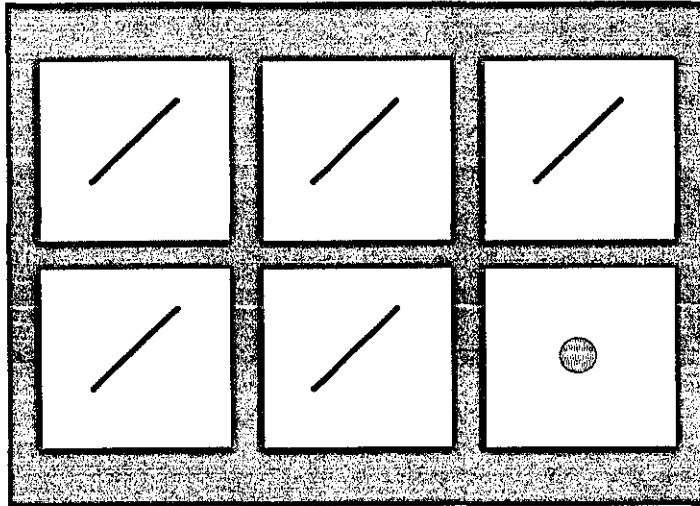
Subtest 'Odd One Out' version 3, item 2
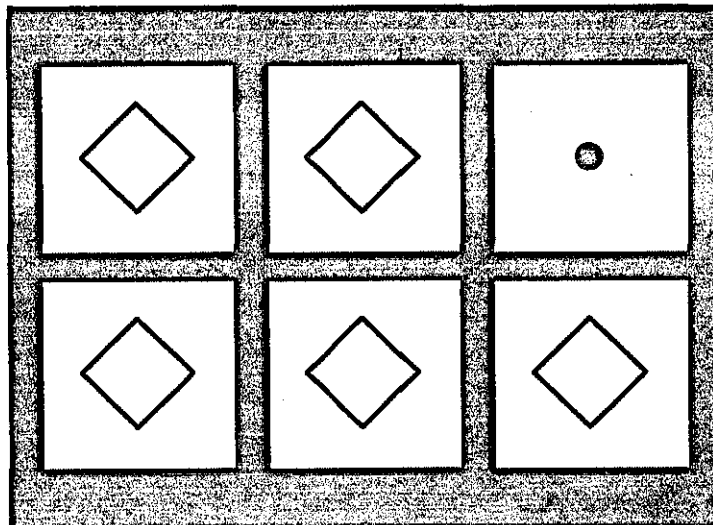
Subtest 'Odd One Out' version 3, item 3
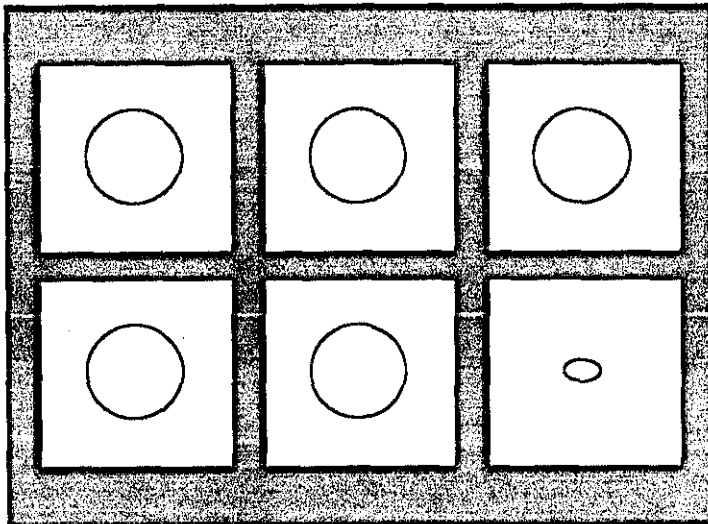
Subtest 'Odd One Out' version 3, item 4
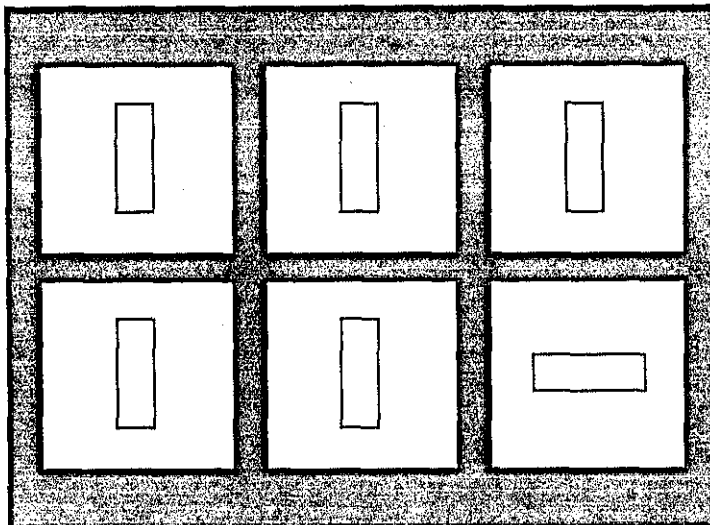
Subtest 'Odd One Out' version 3, item 5
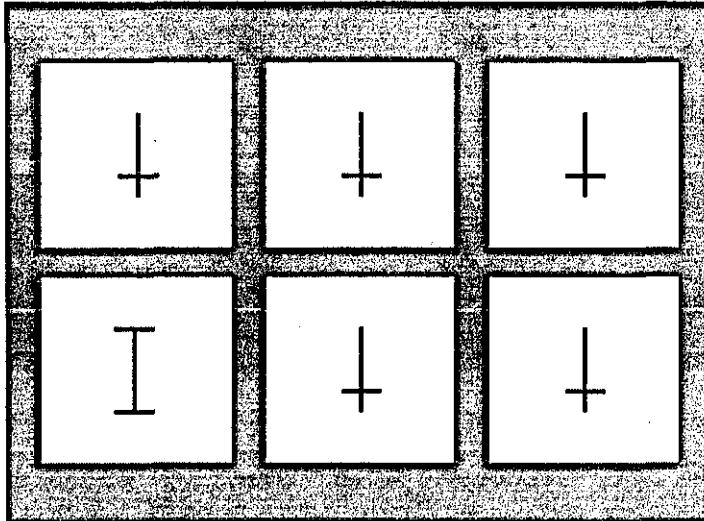
Subtest 'Odd One Out' version 3, item 6
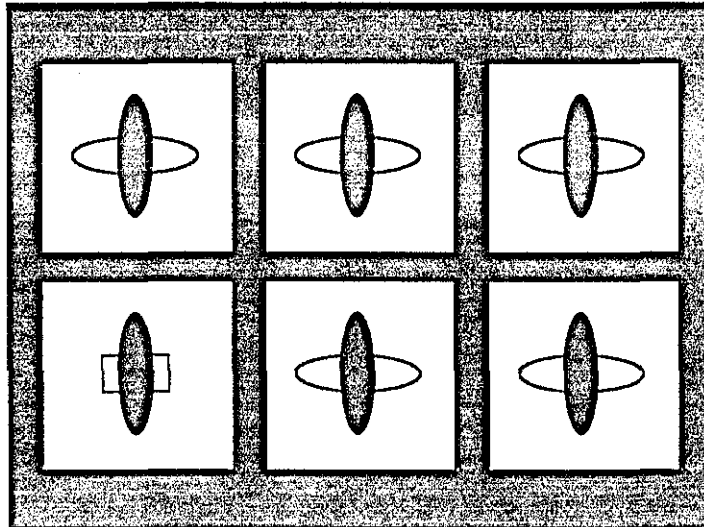
Subtest 'Odd One Out' version 3, item 7
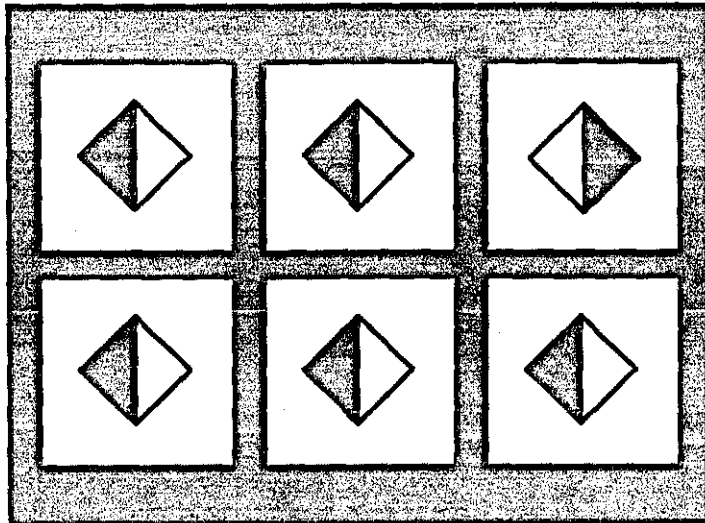
Subtest 'Odd One Out' version 3, item 8
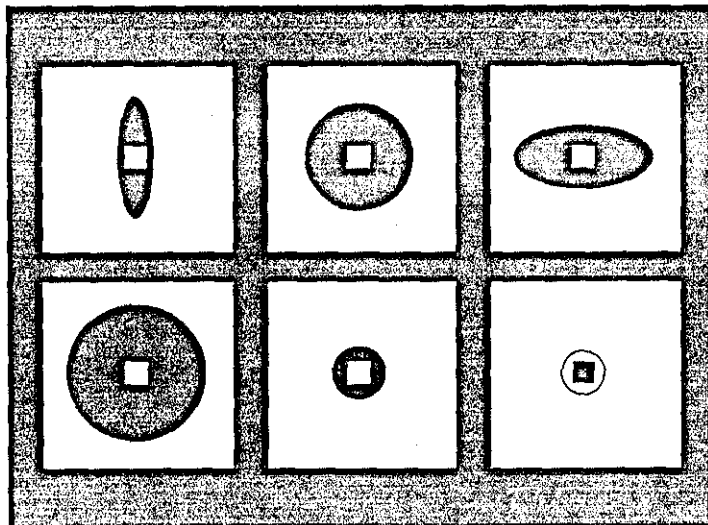
Subtest 'Odd One Out' version 3, item 9
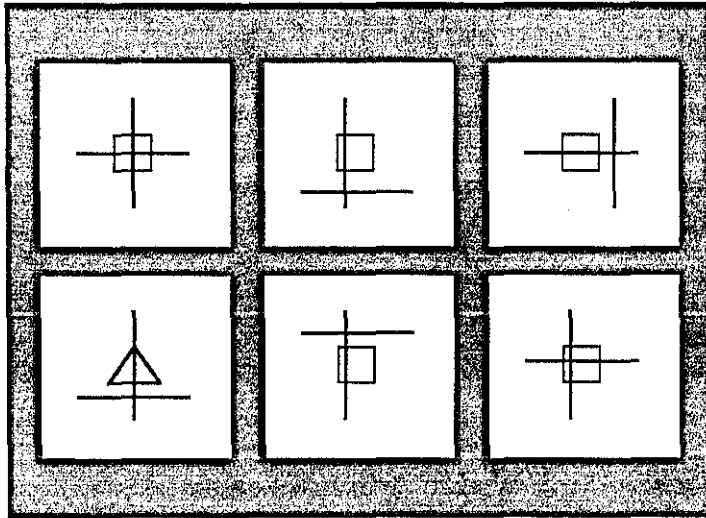
Subtest 'Odd One Out' version 3, item 10
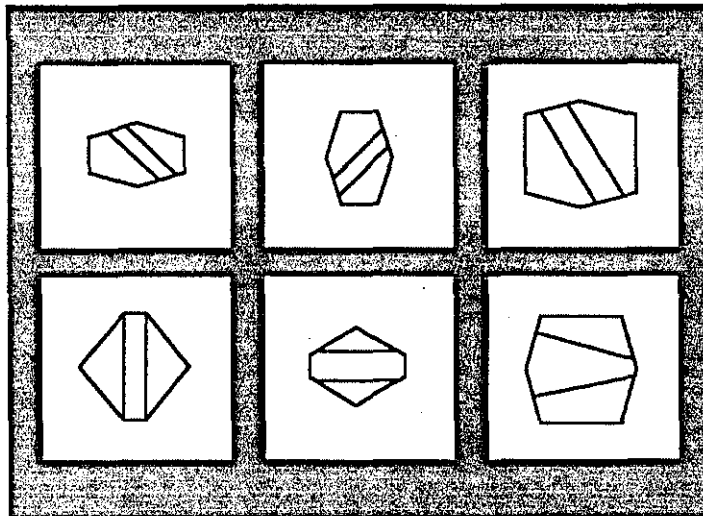
Subtest 'Odd One Out' version 3, item 11
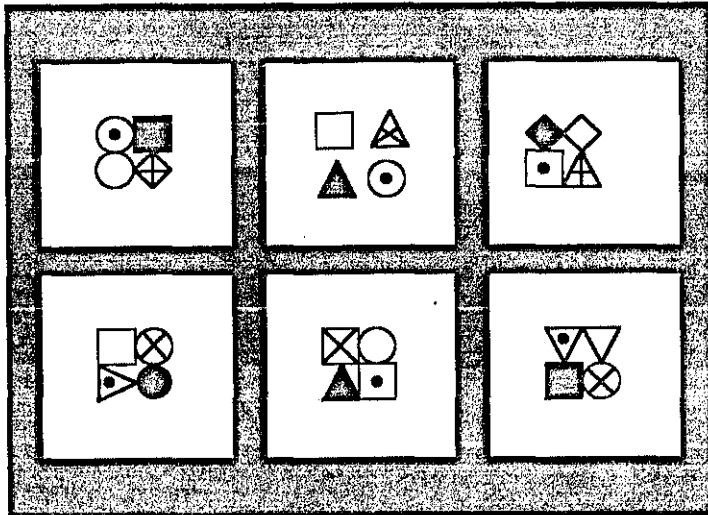
Subtest 'Odd One Out' version 3, item 12
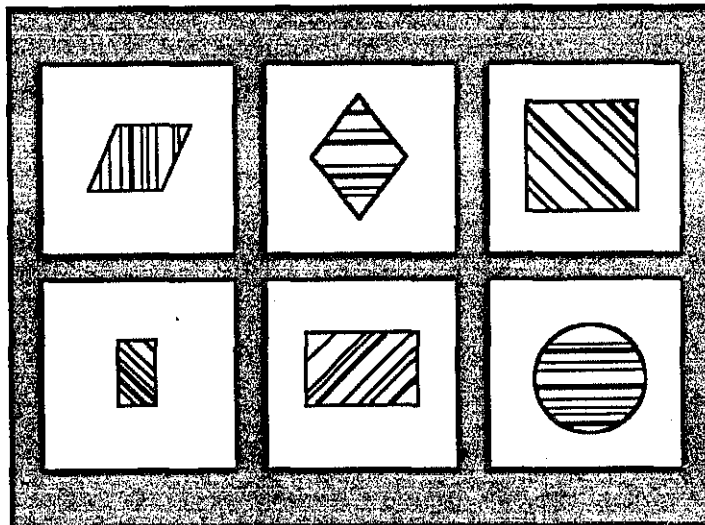
Subtest 'Odd One Out' version 3, item 13

Subtest 'Odd One Out' version 3, item 14
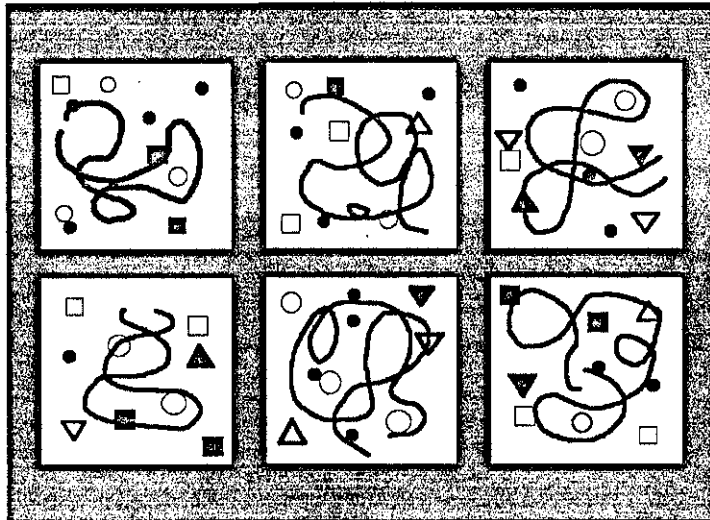
Subtest 'Jigsaw' item 1
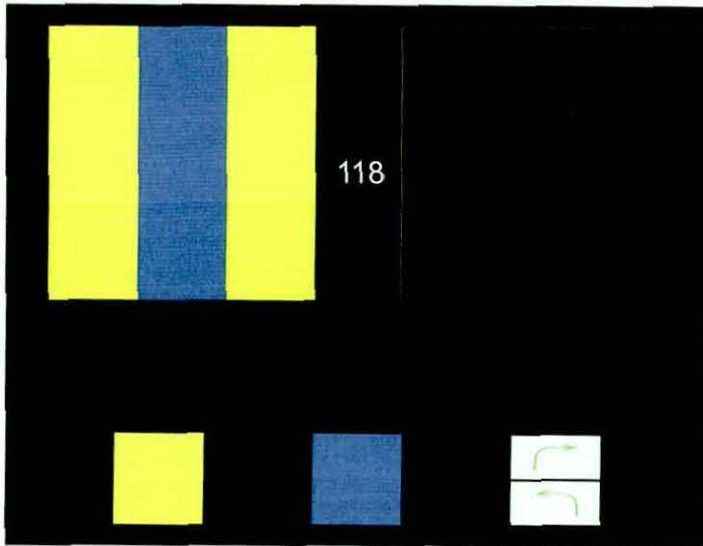
Subtest 'Jigsaw' item 2

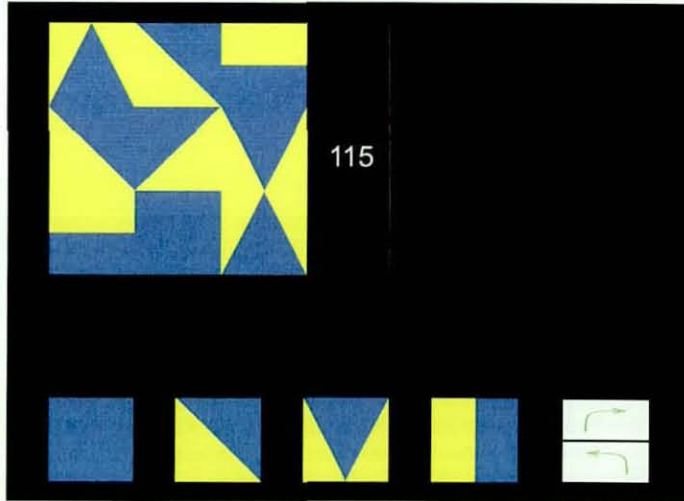Subtest 'Jigsaw' item 3
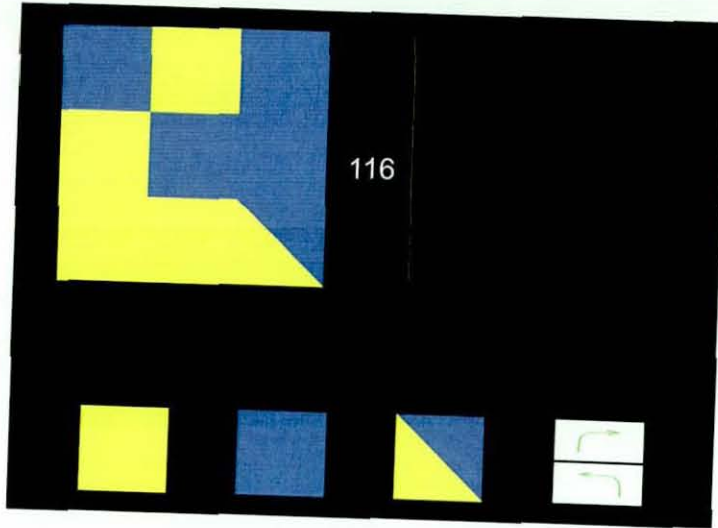
Subtest 'Jigsaw' item 4
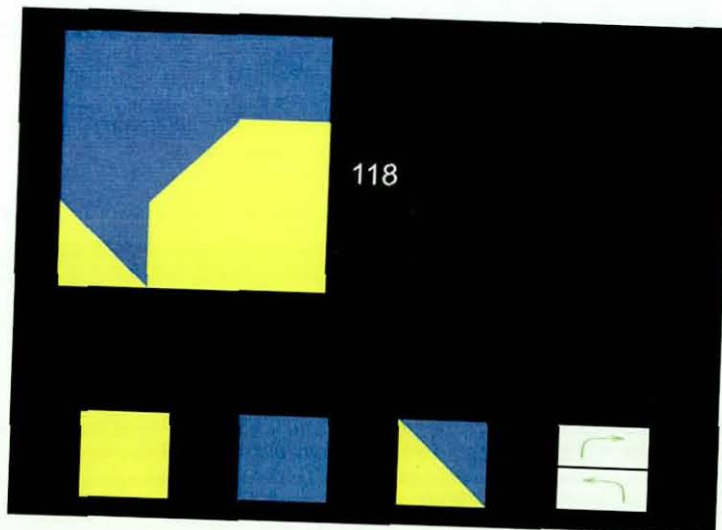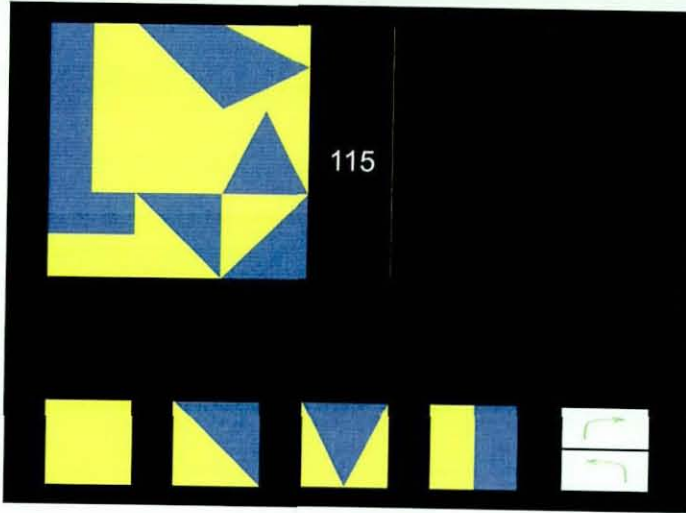
Subtest 'Jigsaw' item 5
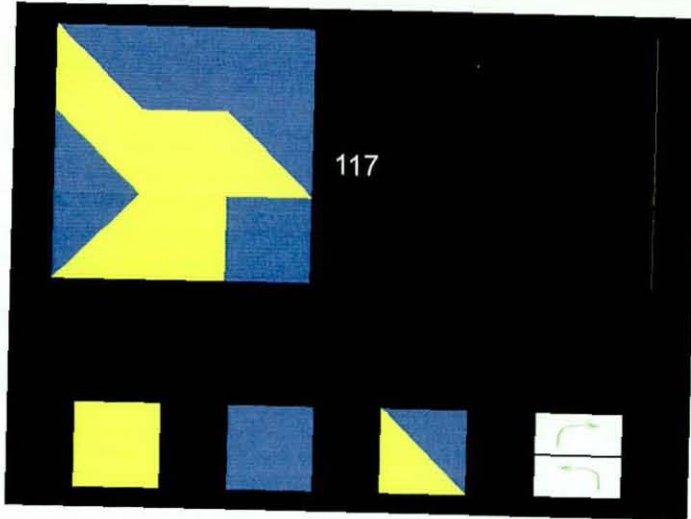
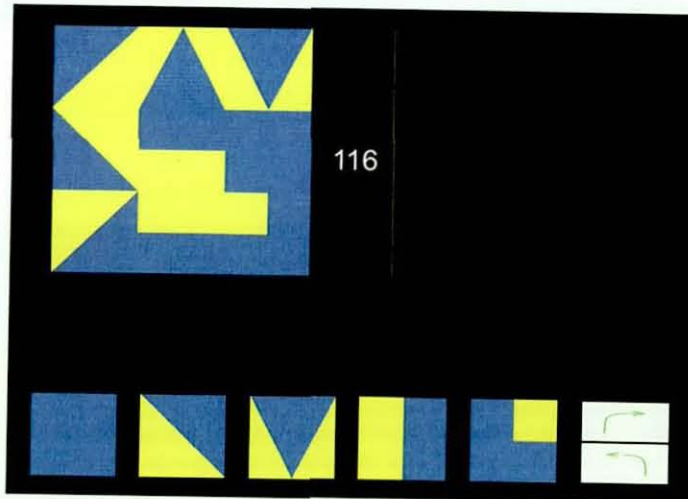Subtest 'Jigsaw' item 6

288
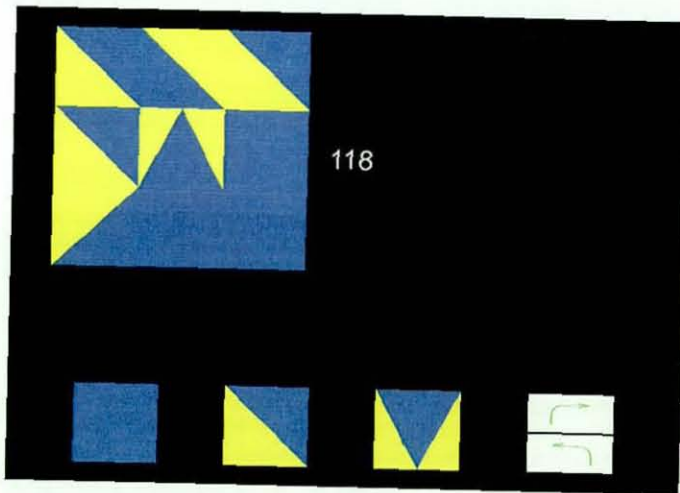
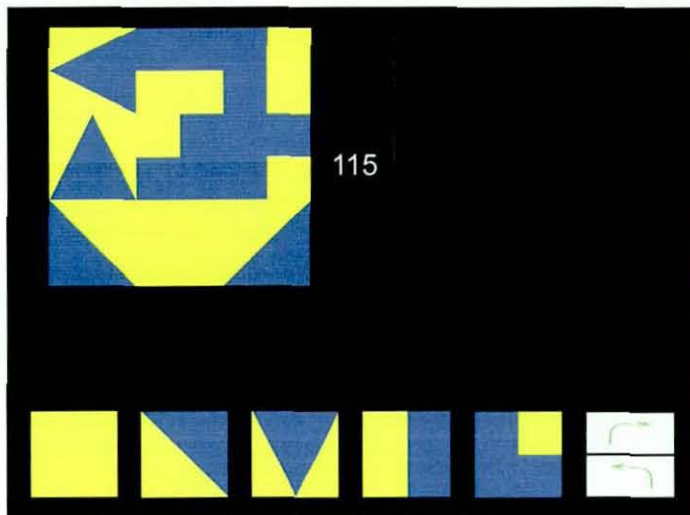Subtest 'Jigsaw' item 7
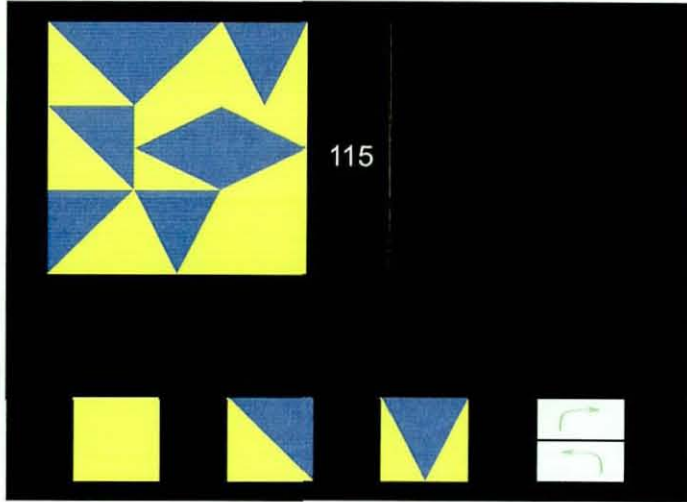
Subtest 'Jigsaw' item 8
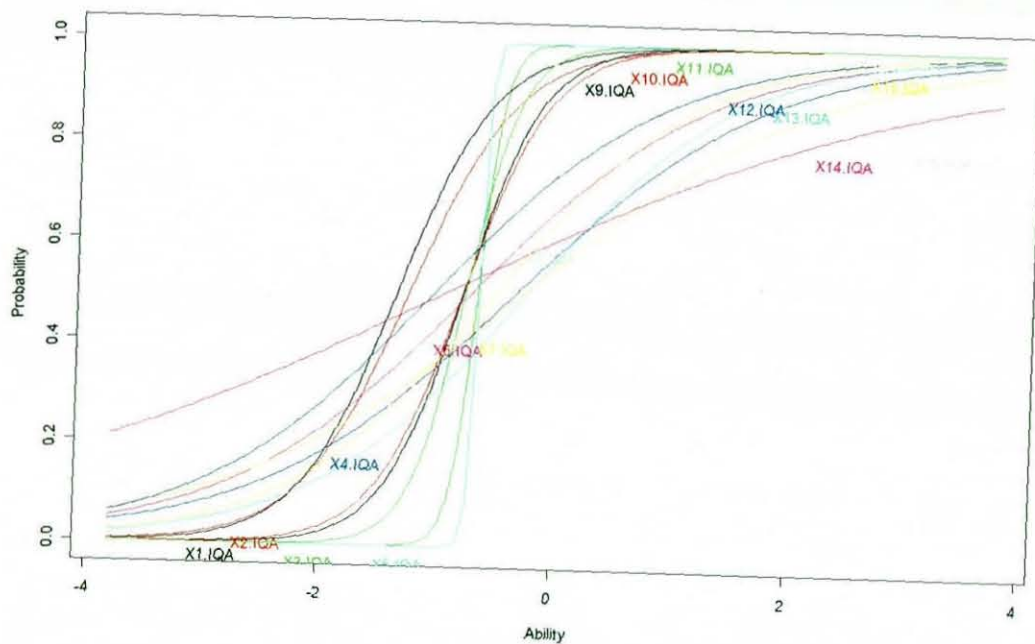

289

Subtest 'Jigsaw' item 9

Subtest 'Jigsaw' item 10

290

Subtest 'Jigsaw' item 11

115

Appendix F: Item characteristic curves for subtest 'Series' version 1

Appendix G: Information letter and consent form for sport-specific studies

GLOBAL GAMES 2009 LIBEREC:

INFORMED CONSENT & ASSENT FORM

Research Project Title

Enhancing Sport for Athletes with Intellectual Disability: Classification Research

Researcher(s)/Organisers
- Jennifer Mactavish (University of Manitoba, Canada), Melanie Gregg (University of Winnipeg, Canada)
- Yves Vanlandewijck (Catholic University of Leuven, Belgium)
- Stephan Bandelow (University of Loughbourough, United Kingdom)
- Jan Burns (Canterbury Christ Church University, United Kingdom)
- Kennet Fröjd (Swedish Development Centre for Disability Sport)

Sponsors INAS-FID, IPC, UK Ministry of Sport

Approved by
- INAS-FID and IPC, as part of the ID-eligibility project action plan
- Ethical committee of the Katholieke Universiteit Leuven
- Ethical Advisory Committee, Loughborough University.
- Education & Nursing Research Ethics Board, University of Manitoba

This consent form, a copy of which you can keep, is part of the informed consent process. It tells you the main idea of the research and what your participation will involve. Please take the read this letter and any other information that comes with it carefully. If you do not understand something, or you want to know about something not mentioned, please feel free to ask. <u>Note</u>: If you have difficulties reading, this letter will be read aloud to you. After each point, you will be asked: *Do you understand what this means? Do you have any questions? Do you agree to this point?*

1. **Purpose of the Research:** *What are we doing and why?*
   The purpose of 2009 Global Games research is to get information that shows how intellectual disability effects sport, and how this might change depending on the specific sport (Athletics, Basketball, Swimming, and Table Tennis). This information is very important because it is required if future opportunities for competing in International Paralympic Committee (IPC) sanctioned events and competitions is to happen.

   For this reason, taking part in the Global Games is very, very important. Being part of the research will not affect your training or competition in Liberec.

   All of the information collected will be stored in a secure data base, which will not be accessible to anyone outside of the research team, and will be used for the purpose of better understanding elite athletes with intellectual disabilities—with the intention of developing systems that meet the requirements of the IPC Classification Code.

2. **Research Procedures:** *What will you be asked to do?*
   <u>All athletes</u> will be asked to complete a computer based battery using a touch screen system. This works a lot like a computer game where images come up on the screen and the athlete touches the screen to indicate his/her choice. This takes about 60 minutes and includes non-verbal items designed to assess skills requiring attention, eye-hand co-

1

ordination, visual (seeing) speed, reaction times and several types of memory. All of these skills are important in sport in general.

We know that some of the skills covered in the computer test can be influenced by sport psychology (goal setting, mental preparation) and coping skills (dealing with pre-competition nervousness). Two questionnaires, only available in English right now, will be used so we will know if this has affected the information from the computer tests. *For only this part of the research*, only athletes from native English speaking countries will be asked to take part. If needed, questions will be read aloud, with the responses recorded with a check mark. It will take approximately 40 minutes to complete the questionnaires.

Besides skills that are important in sport generally, several other areas will be looked at to help us better understand the effect of intellectual disability in different sports and to take into account the effects of training. These include:

(a) In the sport of basketball, athletes will be asked to complete a sport specific series using photographs of common plays to provide information on tactical skills. This test works a lot like the computer based one that all athletes will take, but uses examples that are specific to the game of basketball (approximately 30 minutes). Among other things this will help us to understand the effects of an athlete's intellectual disability in basketball and to account for the effect of training history.

(d) In table tennis, athletes will be asked to do a series of sport specific skill tests (approximately 45 minutes) that will include, for example, serving, service return, and other common game skills.

(c) Video tape (film) recordings of competitions also will be taken for later analysis using standardized technical and tactical observation protocols.

3. Risk Assessment: *Will taking part put me at risk in anyway?*
Taking part in the research will not put the athlete at risk in anyway and will not effect preparation for or competition during the Global Games.

4. Confidentiality: *Who will get to see my information? Will people be able to identify me?*
Complete confidentiality of all records will be maintained. No response will be connected to any individual participant by name. Only the research team will have access to the information, which will be kept in a secured database.

5. Participation and Compensation: *Do I have to do this and what do I get out of it?*
Taking part in the research is completely up to you—it is voluntary. You are free to withdraw at any time for any reason and this will not have any effect on your team membership or ability to compete. You will not get any compensation (for example, gift, prize or money) for taking part but you will have added to a very, very important step in helping to re-open chances for athletes with intellectual disability to compete in IPC games and competitions in the future.

6. Feedback: *What if I have questions or want to know about my results?*
The researcher will be available following data collection to address any questions the athletes may have. If the athletes, coaches and sport organizations are interested in the results of the study they may contact one of the researchers at +3216329127 or by email at debbie.vanbiesen@faber.kuleuven.be. The researcher will then provide a summary of the results by mail or email.

2

GLOBAL GAMES 2009 LIBEREC: INFORMED CONSENT & ASSENT FORM

Signing your name on this form shows that you understand the information about the research, your role and rights as a participant, and that you agree to take part (be a participant). Please show what you are agreeing to take part in by making an "X" in the boxes below:

☐ Computer battery (attention, memory)   ☐ Basketball skills (on computer)

☐ Sport Psychology (questionnaire)   ☐ Table Tennis skill assessment (activity)

☐ Video taping of games/events

By signing you are not giving up your legal rights and not releasing the researchers, sponsors, or involved institutions from their legal and professional responsibilities. You are free, without prejudice or consequence, to stop participating at any time, and you do not have to answer any questions or perform any test items you do not want to. Also, your participation during the project should be as informed as your initial consent, so if you have any questions, or would like further information, please feel free to contact:

Debbie Van Biesen, phone +32 16329127; e-mail: debbie.vanbiesen@faber.kuleuven.be
Katholieke Universiteit Leuven – Faculty of Kinesiology and Rehabilitation Sciences
Department of Rehabilitation Sciences, Tervuursevest 101, B 3001 Leuven, Belgium.

Ethics approval for the research has been granted by: the *Education & Nursing Research Ethics Board at the University of Manitoba* (Canada), the *Ethical Committee of Katholieke Universiteit Leuven* (Belgium), and the *Ethical Advisory Committee of Loughborough University* (United Kingdom).

If you have any concerns or complaints about this project please contact the above-named person. A copy of this consent form has been given to you to keep.

_____
Participant's Signature   Date

_____
Substitute Decision Maker's Signature   Date

_____
Please indicate the legal relationship by which power to consent has been delegated

_____
Researcher and/or Delegate's Signature   Date

I would like a copy of my assessment.	☐ Yes	☐ No
Print name and Mailing address:_____		
_____		
_____		

3