Comparing Individuals with Learning Disability and those with Borderline IQ: A Confirmatory Factor Analysis of the Wechsler Adult Intelligence Scale (3rd Edition).

Hannah MacLean

A thesis submitted in partial fulfilment of the requirements of the University of Edinburgh for the degree of Doctor of Clinical Psychology

2010

Abstract

Background: Support for the four factor construct validity of the third edition of the Wechsler Adult Intelligence Scale (WAIS-III) has been found in clinical and non clinical populations but some studies question whether more complex models consistent with the concepts of fluid and crystallised intelligence provide a better explanation of the data. The WAIS-III is frequently used in the diagnosis of learning disability, however, previous exploratory factor analysis of data from a population with low IQ did not support the explicit four factor structure of the WAIS-III.

Method: A confirmatory factor analysis of the WAIS-III was carried out on data from people with severe and significant learning disability and people with borderline IQ (IQ = 70-79).

Results: The data from the borderline IQ sample and the sample with significant learning disability showed at best a weak fit to the explicit four factor models and more complex five or six factor models. However fit of the data from the sample with severe learning disability was poor for all models.

Discussion: The findings show little support for the explicit four factor construct validity of the WAIS-III for people with borderline IQ or significant or severe intellectual impairment. Some support is found for the direction taken by the new Wechsler children's and adult scales (WISC-IV & WAIS-IV) in aligning interpretation of the scales more closely to concepts such as fluid and crystallised theory. The research also suggests the cut-off point of IQ 70 is not reflective of an actual difference in cognitive profile as measured by the WAIS-III. Limitations of this study and implications for further research are also discussed.

Acknowledgements

With thanks to Dr. Karen McKenzie for her advice and encouragement. To Dr Matthias Schwannauer and Dr. Gill Kidd for their comments. And to my family for their love and support.

Word Count: 25885

D. Clin. Psychol. Declaration of own work

This sheet must be filled in (each box ticked to show that the condition has been met), signed and dated, and included with all assessments - work will not be marked unless this is done

Name: Hannah MacLean

Assessed work: Thesis

Title of work: Comparing individuals with Learning Disability and Borderline IQ: A confirmatory factor analysis of the Wechsler Adult Intelligence Scale (3rd Edition).

I confirm that all this work is my own except where indicated, and that I have:

- Read and understood the Plagiarism Rules and Regulations in the Programme Handbook ☐
- Composed and undertaken the work myself ☐
- Clearly referenced/listed all sources as appropriate ☐
- Referenced and put in inverted commas any quoted text of more than three words (from books, web, etc) ☐
- Given the sources of all pictures, data etc. that are not my own ☐
- Not made undue use of essay(s) of any other student(s) either past or present (or where used, this has been referenced appropriately) ☐
- Not sought or used the help of any external professional agencies for the work (or where used, this has been referenced appropriately) ☐
- Not submitted the work for any other degree or professional qualification except as specified ☐
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources) ☐
- Complied with other plagiarism criteria specified in the Programme Handbook☐
- I understand that any false claim for this work will be penalised in accordance withthe University regulations ☐

Signature …………………………………………… **Date** …………

Table of Contents

List of Tables Page

List of Figures Page

List of Appendices

1. Introduction

Intelligence is central to the diagnosis of learning disability. However, there are particular measurement, cultural and social issues involved in assessing intelligence in people with learning disability and some authors have concluded that the diagnostic criteria of significant impairment in intellectual functioning (IQ less than 70) in learning disability is flawed, socially constructed, atheoretical and arbitrary. Only limited research has been carried out to examine the intellectual profiles of people with learning disability and low intellectual functioning. Some of this research suggests that the intellectual profile of these groups, when measured by the popular Wechsler intelligence scales, does differ from that of the general population. This has a number of potential implications in relation to how IQ is conceptualised and measured in people with learning disability and also suggests that the IQ cut-off may not be arbitrary and instead may have a theoretical underpinning. The present study aims to examine if the factor structure of IQ, as measured by the WAIS III, differs for people with learning disability and people with borderline intellectual functioning.

The introduction will begin by providing a brief overview of learning disability and follow with an outline of intelligence and its measurement with a particular focus on the Wechsler scales. The relevance of these scales will be explored in relation to diagnosis of learning disability followed by a discussion of their validity for people with low IQ. The introduction will end by outlining the rationale and aims of the study.

1.1 Learning Disabilities – definitions and diagnosis

Changing social, cultural and professional trends often dictate popular terms for people with a learning disability (American Association on Mental Retardation (AAMR), 2002). These terms often serve a purpose such as distinguishing people in terms of capacity or entitlement. As social views change, terms can become outdated and offensive; as was the case with the language of the early 1900s, when terms such as *idiot, imbecile and moron* were used to describe individuals falling within particular IQ bands (Digby, 1996). Currently, in Britain the term *learning difficulty* is sometimes considered less pejorative (British Institute of Learning Disability, 2004). However, this term can be confusing for professionals, particularly as individuals who have problems in more discrete academic domains, such as those with dyslexia, are clinically described as having a difficulty rather than a disability (Royal College of Psychiatrists, 2009). *Learning disability* is widely used in the UK however, around the world, terms such as *mental retardation* and *intellectual disability* are often used interchangeably (AAMD, 2002). As such, differences in terminology can make replication and comparison of research more difficult. For the purposes of this paper, *learning disability* will be used to describe people who meet the British definition of a learning disability.

The three core criteria for diagnosis of a learning disability are:

1. Significant impairment of intellectual functioning
2. Significant impairment of adaptive/social functioning
3. Age of onset before adulthood (British Psychological Society,(BPS), 2002).

In the white paper: Valuing People: A new Strategy for Learning Disabilities for the 21st Century (Department of Health (DoH), 2001) learning disabilities are described at four levels: mild, moderate, severe and profound. Figures for Scotland suggest 20 people for every 1,000 have mild or moderate learning disability and 3 to 4 people for every 1,000 have profound or multiple disabilities. This equates to around 120,000 people in Scotland with learning disability (Scottish Executive, 2000).

The British Psychological Society (2002) uses the terms 'significant' and 'severe' to describe levels of learning disability. In terms of intellectual functioning, IQ scores falling between 55-69 are considered to represent significant intellectual impairment and those falling below 55 are considered to represent severe intellectual impairment. When an individual requires intermittent or limited support, their adaptive functioning is described as significantly impaired. When support required is extensive or pervasive, impairment is described as severe. Given an onset of difficulties before age 18 a diagnosis can be made and the level of learning disability can be established.

The first criterion required to diagnose learning disability is significant impairment of intellectual functioning and typically has been described with reference to IQ scores gained from objective measures, such as the Wechsler Intelligence scales (Psychological Corporation, 1981; 1997; 2008). These scales assume performance will be normally distributed across the population with the majority around the average IQ of 100 and ninety-five per cent of the population falling within two standard deviations of the mean. Above 100, intellectual functioning can be defined as high average, superior and very superior, respectively, reflecting the increasing deviation from the mean. Below 100, increasingly poorer intellectual functioning can be defined as low average, borderline, extremely low, respectively (Lichtenberger & Kaufman, 2009). According to the BPS (2002), those with at least significant impairment in intellectual functioning fall below the cut-off point of IQ 70 and meet one criterion for a diagnosis of learning disability.

This cut-off reflects the point at which individuals' scores are more than two standard deviations from the mean IQ for the general population. For those with an IQ above 70, the learning disability term would not be considered unless the score fell within the confidence interval for IQ 70. An actual score on a test gives an indication that the true IQ is within a small range of points which may actually straddle the cut-off point of 70. This range of scores takes into account inherent error in the test and therefore when confidence levels are applied to an IQ slightly over 70 the lower limit may fall under the cut-off point and allow the criterion for intellectual functioning at the significant impairment level to be met. However, when the criteria for the diagnosis are applied strictly to determine qualification for support and services, it may be the case that an individual with an IQ score of 69 who also meets the other diagnostic criteria receives support and services not available to a similar individual with a borderline IQ of 70. Therefore the tests used to assess intellectual functioning and how they are interpreted are of great significance (Lin, 2003).

The second criterion for diagnosis of learning disability focuses on how an individual functions in day to day life. Like some measures of intelligence, tests of adaptive functioning are criticised for lack of cultural specificity (Winters *et al.*, 2005), however, standardised tests of functioning, completed by someone who knows the individual well, can provide a more ecologically relevant assessment. The Diagnostic and Statistical Manual – 4th Edition (DSM-IV; American Psychiatric Association, 1994) define a significant impairment of adaptive functioning in terms of concurrent deficits or impairments in at least two areas such as communication, work, social /interpersonal skills, home-living, self care, use of community resources, self direction, functional academic skills, work, leisure, health and safety. In the UK, adaptive functioning is often assessed by psychologists using tests such as the Vineland Adaptive Functioning Scale (Sparrow *et al.* 1984) or the Adaptive Behaviour Scale (ABAS; Harrison & Oakland, 2000). Both are standardised tests using a semi-structured interview format which are completed by someone who knows the person well. The tests are also normed for people with a learning disability and have been recently updated: Vineland Adaptive Behaviour Scale – Second Edition (Vineland II; Sparrow, *et al.* 2005) and Adaptive Behaviour Assessment System - Second Edition (ABAS-II; Harrison & Oakland, 2003).

Despite increased attention on adaptive functioning, the Diagnostic and Statistical Manual 4th Edition – Text Revision (DSM-IV-TR), the current diagnostic publication from the American Psychiatric Association (APA), suggests IQ is the key component of mental retardation[1] (Fitzgerald *et al.* 2007). IQ is often used synonymously with the term intelligence and while traditionally IQ has been more simply defined, the concept of intelligence has been more difficult to operationalise.

[1] Term 'mental retardation' is used by the APA but is not used commonly in the UK. For the purposes of this paper the term *learning disability* is used rather than mental retardation.

1.2 Defining intelligence

Early in the twentieth century the term *intelligence* was causing great frustration even to those involved in its study: '…a mere vocal sound, a word with so many meanings that finally it had none.' (Spearman,1927 in Reber,1985, p.379). Key abilities identified as components of intelligence included reasoning, judgement, insight, abstraction and dealing with new situations and were contrasted with ideas about a single general factor. Boring (1923) famously proposed that "intelligence is what intelligence tests test" (Sternberg, 2000 p.7) in an attempt to stimulate further discussion and refinement of a definition.

The editors of the Journal of Educational Psychology studied the definitions of intelligence provided by contributors to the journal such as Thorndike, Terman and Colvin (Sternberg, 2000). These experts' definitions included ideas about abstract thinking, adaptation to the environment as well as more concrete mental constructs such as memory, imagination, judgement and reasoning. Sixty-five years later, a follow-up study of twenty-four experts demonstrated there was still some agreement with some of the earlier definitions, with emphasis on reasoning, problem solving, decision making and adaptation to the environment. There was new emphasis on culture and context as well as ideas about meta-cognition and executive functions, however, experts continued to be divided in their views of intelligence as a single entity or a sum of many abilities (Sternberg, 2000).

The development of testing procedures while helping refine the term, still demonstrate how most notions of intelligence are inextricably tied to their means of measurement (Reber, 1985). For example, intelligence defined as successful learning in school is measured by tests predicting academic success which measure behaviours equated with successful learning. This circularity does not move definitions much further forward from Boring's (1923) well documented attempt. The endeavour is further confounded by the finding that, in tests of intelligence, success is determined by the priorities of those who design the test and whichever social and cultural parameters the designers subscribe to.

"Ultimately *intelligence* will be, conceptually, what it has always been, *the ability to profit from experience* and pragmatically, what it has become, that which the intelligence tests measure" (Reber, 1985, pp.379-380).

The following section will look at key theoretical models of intelligence and their potential impact on defining learning disability. A more comprehensive description of theories of intelligence can be found in Davidson and Downing, (2000).

1.3 Theories of Intelligence

Intelligence has been described in terms of implicit and explicit theories (Sternberg, 1997; Brody, 2000). Implicit theories or folk theories can be important in how individuals judge themselves and others. They may not be informed by theory but they are crucial to how individuals make judgements in their own lives. They may be culturally and developmentally specific and may provide the basis of more formal explicit theories. They may also highlight the weaknesses of existing formal models by highlighting discrepancies between these and lay experiences (Sternberg, 2000).

Expert or explicit models of intelligence have been informed by several different metaphors (Sternberg, 1985). These include a geographic metaphor with intelligence viewed as a map of the mind; a computational metaphor which looks at information processing; an anthropological metaphor which looks at the individual's relationship to his/her cultural context; and others such as the biological, genetic-epistemological, systems and sociological. Despite their differences the evaluation of each of these ideas has added to the ongoing debate about the nature of intelligence and how it can be defined and tested (Sternberg, 2000).

Early work by researchers such as Spearman (1904) used performance on mental tests as a way of measuring intelligence and exploring the possible structure of intelligence. Statistical tools such as factor analysis have been helpful in developing these structural models by exploring correlations in performance on mental tasks. For example, if performance on one task is highly related to performance on another task, this suggests there is a common factor which is used by both tasks. Spearman's analyses (1927) precipitated the use of the term *g* to explain a single most important factor used for carrying out mental tests. It was perceived as a general factor that all mental activity would tap into. More specific but less important first order factors *s* were identified for particular tasks. Thurstone, (1938) also used factor analysis to examine the structure of intelligence but did not find support for a general factor and instead found evidence of seven independent factors: verbal comprehension, word fluency, space, number facility, perceptual speed, induction and memory.

To address the discrepancy between Spearman's and Thurstone's findings, subsequent theories have adopted a hierarchical structure which suggests the presence of a general factor at the apex of the hierarchy with lower level factors at the bottom of the hierarchy (Davidson & Downing, 2000). One such theory combines the idea of fluid intelligence and crystallised intelligence (Cattell, 1963). Crystallised intelligence (*gc*) tasks were those considered to be dependent on education and experience whereas fluid intelligence tasks (*gf*) were thought to be linked to the efficiency of internal mechanisms. This distinction was thought by Cattell to reflect a genetic versus environmental divide between the two types of intelligence. A revised version of *gf-gc* theory (Horn, 1986) is a two-stratum hierarchical model with more than forty first order factors, including those factors specified by Thurstone (1938) as primary mental abilities. It suggests both *gf* and *gc* have environmental and biological bases involving low and high level functioning of internal mechanisms. The theory is also supported by studies across the life span which suggests some abilities remain stable (e.g. long term memories) while others deteriorate (e.g. processing speed and short term memory; Davidson & Downing, 2000).

1.3.1 Three Stratum Theory

Carroll's Three Stratum Theory (1993) can be viewed as a pyramid with a factor similar to *g* at its apex. This factor is considered to have a high genetic component and, like Spearman, Carroll viewed *g* as the basis of all mental activity. The second stratum contains eight second-order factors, similar to those in *gf-gc* theory, which are correlated to *g* to different degrees. More specific abilities are reflected in factors at stratum one. While three strata are defined it is acknowledged there may be more intermediary levels which better describe particular abilities (Davidson & Downing, 2000).

1.3.2 Cattel-Horn-Carroll (CHC) Theory

In 1993 Carroll published *Human cognitive abilities: A survey of factor-analytic studies*. The book summarised more than 460 different factor analytic studies of human cognitive abilities. Jenson (2004) describes the magnitude of the work carried out by Carroll (from McGrew, 2003):

"Carroll's magnum opus thus distills and synthesizes the results of a century of factor analyses of mental tests. It is virtually the grand finale of the era of psychometric description and taxonomy of human cognitive abilities. It is unlikely that his monumental feat will ever be attempted again by anyone, or that it could be much improved on. It will long be the key reference point and a solid foundation for the explanatory era of differential psychology that we now see burgeoning in genetics and the brain sciences (p.5).

Carroll suggested the Horn-Catell *gf-gc* model provided the best evidenced and most acceptable theory of the structure of cognitive abilities. Thus, the writings of both Horn and Carroll were then synthesised to create CHC theory. The theory was adopted into the field of applied intelligence testing and became the predominant theory in which to base and evaluate modern intelligence tests (McGrew, 2003).

CHC theory is a hierarchical framework with three strata of cognitive abilities. Despite the difference in opinion between Horn and Carroll on the existence of *g*, the model adopts Carroll's stance and stratum III refers to *g* or general intelligence; stratum II refers to ten broad cognitive abilities, Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-term Memory (*Gsm*), Visual Processing (*Gv*), Auditory Processing (*Ga*), Long-term Retrieval (*Glr*), Processing Speed (*Gs*), and Decision/Reaction Time or Speed (*Gt*), Reading and Writing (*Grw*), and Quantitative Knowledge (*Gq*; McGrew & Flanagan, 1998). These broad cognitive abilities overarch approximately 70 narrow cognitive abilities.

Proponents such as Flanagan and Kaufman (2004) argue CHC theory has succeeded in bridging the gap between theory and practice. Projects such as the Carroll Human Cognitive Abilities Project (McGrew, 2003) celebrate this long history of factor analytic studies and aim to extend the work by re-analysing historical data sets, confirming results using more recent statistical tools and carrying out prospective analysis of current data.

These theories fit well with a neuropsychological approach to understanding learning disability. Standardised tests designed to tap into particular neuropsychological function can be normed so that an individual's performance can be linked to typical performance for someone of a similar age. Provided an accurate picture of typical performance can be determined, atypical performance can also be highlighted.

1.3.3 Contextual Models

The Ecological Model (Berry & Irvine, 1986) also takes a hierarchical approach but focuses on the contexts in which an individual may experience success rather than focusing on abstract mental tasks. At level one, intelligence can be defined by successful interaction with the physical environment in order to acquire customs and achieve goals. At level two, the individual can succeed in repeating experiences and training to develop socially and culturally appropriate skills, traits and attitudes. At level three, an individual can react to short term demands in their immediate environment. At level four the individual can act or behave in an artificial environment for the purposes of assessment. While highly structured testing environments improve the internal construct validity and reliability of a test, they lack ecological validity; the extent to which the results can be generalised to other situations. Berry (1994) explains how an absolutist view of intelligence assumes that the mental abilities and testing procedures relevant to one society are relevant to another. In comparison, contextual models emphasise the dynamic nature of intelligence and the need for it to be assessed with reference to relevant contexts (Davidson & Downing, 2000). With this approach an individual whose performance was poor on mental tasks may still experience success by interacting successfully with their own environment and developing skills appropriate to that environment.

1.3.4 Complex systems models

Complex systems models also conceptualise intelligence as a dynamic rather than static force. They acknowledge the important role of context, physiology and cognitive factors. One such example is the Triarchic Theory of Intelligence (Sternberg, 1985, 1988, 1997). The internal aspect of intelligence refers to the lower order mental processes which are required for successful problem solving. Higher order processes or meta-components can be used as guides to problem solving and according to Sternberg (1985) explain why *g* is common to many factor analytic studies (Davidson & Downing, 2000). The external aspect of intelligence refers to our adaptation to a situation and ability to apply our internal intelligence to a variety of contexts. The experiential aspect of intelligence allows us to use our existing knowledge and experience to solve new problems or process new information automatically. The model allows for strengths in the three aspects of intelligence to be differentiated and suggests a common feature of intelligent individuals across cultures is their ability to make the most of their strengths while compensating for their weaknesses. Similar to the contextual models, the complex systems model does not rely on performance on individual mental tasks to define intelligence. As such it provides arguably a more ecologically valid approach to the measurement of intelligence but one which is difficult to measure objectively.

1.3.5 Multiple Systems Approach

The multiple systems approach (Gardner, 1983, 1998) suggests intelligence can only make sense when embedded in a context natural to the individual. As such, evidence for Gardner's theory comes from more naturalistic settings and assessments and is more difficult to generalise and replicate. Gardner's eight intelligences include three abilities often tested in traditional intelligence tests: linguistic, logical-mathematical, spatial and five others: musical, bodily-kinaesthetic, intrapersonal, interpersonal and naturalistic and are developed through a combination of genetic influence, socialisation and training.

Summary

Intelligence theories can broadly be divided into two domains. One domain favours internal validity and the focus on intelligence as performance on individual mental tests which are completed under tightly controlled conditions and allow an individual's performance to be ranked and compared to other individuals. The assumption is that these abilities underlie intelligence between and within all populations. The other domain focuses more on individuals' successes in their own context and their ability to adapt to new environments. In the field of applied psychology, and in particular clinical psychology, intelligence testing along the former lines has been favoured. This can be seen by the type of tests which are most commonly used amongst practitioners. These tests have their foundations in strong internal validity verified by statistical techniques such as factor analysis. The following section will look at the history and development of intelligence testing and its relevance to the field of learning disability.

2. Development of Intelligence Testing

Intelligence testing has been used extensively to understand individual differences in intellect. While its usefulness in this regard is beyond question it is also clear that testing has at times been used to support implicit theories about intelligence and society. Sternberg (2000) describes what he calls the Hamiltonian view. This view suggests that people are born with different levels of intelligence and the high IQ elite need to take care of the low IQ masses, unable to take care of themselves. Books such as 'The Bell Curve' by Hernstein and Murray (1994) indicate there is ongoing interest by some in delineating intelligence along elitist lines. Sternberg's own view (Sternberg, 2004) is similar to what he describes as a Jeffersonian view:

> "people are equal in terms of political and social rights and should have equal opportunities, but they do not necessarily avail themselves equally of these opportunities and are not necessarily equally rewarded for their accomplishments...In this view, the goal of education is not to favour or foster elite, as in the Hamiltonian tradition, but rather to allow children the opportunities to make full use of the skills they have." (Sternberg, p.12).

In the early Twentieth century, segregation according to intellectual ability was aligned with the political agenda of the time and Binet (1905) created a means of examining the differences between individuals by looking at memory, reasoning and judgement. The tests resulted in the segregation of children with learning disability in the classroom. This was replicated in North America and the Stanford-Binet test

(Terman, 1916) became an adapted version of the Binet test made relevant for the United States, using an American standardisation sample. Terman also became known for introducing the idea of IQ as mental age divided by chronological age and multiplied by 100.

In the late nineteenth century, Galton's (1869) study of sensory abilities failed to demonstrate any predictive validity for more objective measures of intelligence such as academic achievement. However, his work did initiate the interest in correlations between cognitive abilities and the importance of test reliability; the idea that a person completing a test on one occasion would be expected to perform in a similar manner on a future occasion. The idea of test validity was also introduced; that is the extent to which a test measures what it claims to measure. With the start of World War I, there was a need for valid and reliable tests to facilitate recruitment and appropriate placement into the army. The Army Alpha Test (Yerkes, 1921) was similar to the Stanford-Binet (Terman, 1916) while the Alpha Beta (Yerkes, 1921) used standardised non verbal tests. Over one and a half million men were tested (Lichtenberger & Kaufman, 2009) An additional test, the Army Performance Scale Examination allowed testers to assess a further dimension:

"to prove conclusively that a man was weakminded and not merely indifferent or malingering" (Yoakum & Yerkes, 1920 in Lichtenberger & Kaufman, 2009, p.4).

2.1 Validity of intelligence tests

In order for intelligence tests to be considered valid tests of intelligence they would be expected to demonstrate good predictive ability between one test and another measure of intelligence. Good internal construct validity [2] would also be important and this would be shown by understanding how variance in the test scores could be accounted for by different key concepts thought to relate to intelligence. Statistical procedures such as factor analysis help to extract key elements from complex data sets by highlighting the correlations between variables. In the field of intelligence, factor analytic work by Spearman (1904) and Thurstone (1938) helped others formulate models of intelligence which could take account of a general intelligence factor and specify other factors more directly involved in individual information processing tasks. Kline (1994) describes a factor as a:

"construct or dimension which can account for the relationships (correlations) between variables." p.13.

In an exploratory factor analysis the data is correlated and the resultant loadings are assessed for significance: the higher the loadings the more convincing the factor. At a minimum, a variable is considered to load onto a factor if the correlation is at least 0.32 (Tabachnick & Fidell, 2001), while a higher factor loading of 0.63 would be considered very good (Comrey & Lee, 1992). In this process there are no theoretical constraints on the data and the results generate hypotheses about the dependency or independency of relationships between factors; thus the process is usually undertaken early on in research (Tabachnick & Fidell, 2001). Later in the research process, a confirmatory factor analysis (CFA; discussed in detail in the Methodology section) can test these hypotheses for theoretical fit. A target matrix is set based on hypotheses, and the actual correlation matrix is checked against the target matrix for

[2] Construct validity is defined as the extent to which underlying traits the test claims to measure are actually being measured (Reber, 1995).

goodness of fit. In this way it can be seen how much of the variance is accounted for by the proposed factors (Kline, 1994). Kline suggests:

> "the fact that a latent structure has been confirmed does not imply that the structure is what one was looking for or is of psychological importance. As with all factors these must be identified not just from their loadings (merely face validity) but with reference to external criteria or location in factor space." (Kline, 1994, p72).

Factor analysis of tests such as the WAIS-III help us understand intelligence in a particular framework, however, it may not be the only framework helpful in understanding intelligence in any given situation. For example, hierarchical theories developed out of early factor analytic studies and, while internal validity and reliability of the tests were good, debate was stimulated about the ecological validity of static intelligence tested under artificial circumstances (Gardner, 1998). Contemporary theories including complex systems theories, just as the traditional models place emphasis on the importance of adaptability, have also sought to address ecological validity by giving context more prominence and broadening the concept of intelligence (Davidson &Downing, 2000). However, these theories which are less experimentally specified than traditional models (Davidson & Downing, 2000) perhaps lend themselves less well to stringent testing protocols and encourage the use of more rigorous tests which provide easily quantifiable data such as IQ. The psychometric properties of scales such as the Wechsler scales allow recruiters, including psychologists, researchers and employers, to be confident of the quality of a test.

In the measurement of intelligence, the Wechsler scales are amongst the most popular for this reason (Camara *et al*. 2000). They can hold a powerful place in legal settings where an objective measure of intelligence is required to assess eligibility for services and support, despite the broader assessment approach advocated by the definition of a learning disability. The following section will take a closer look at the Wechsler scales.

2.2 History of the Wechsler scales

David Wechsler used components of the Stanford-Binet, the Army Alpha test and the Army Performance Scale Examination to create the Wechsler-Bellevue scale (1939). It was designed to be a standardised test able to measure intelligence distinct from verbal skills in adolescents and adults. Wechsler had worked in the recruitment of armed forces from 1917 and later in the Bellevue Hospital, a psychiatric institution in the United States, which helped him to understand how heavily dependent the current tests were on good verbal skills. The old tests disadvantaged non English speakers or those who were illiterate. Wechsler also wanted his new design to reflect his clinical perception that verbal abilities were distinct from performance abilities and as such he ascribed equal weighting to the Verbal and Performance scales in calculating an overall intelligence score. He felt that IQ testing should be a window to an individual's personality and used others' tests to create a test battery consistent with his own clinical and practical experience rather than focusing on a theoretical foundation (Lichtenberger & Kaufman, 2009). Wechsler (1944) defined intelligence as, "The aggregate of the global capacity to act purposefully, think rationally, to deal effectively with the environment", (in Gross, 1996, p.710) and subscribed to the view that intelligence reflected the sum of many different abilities. As such, the Wechsler-Bellevue subtests were designed to measure different abilities while correlating with each other in order to give a reflection of an overall *g* factor. He also acknowledged the influence of motivation, persistence and goal awareness on performance, suggesting that a significant amount of variation in scores could be affected by these factors. The Wechsler-Bellevue scale was published in 1939 by The Psychological Corporation after Wechsler had standardised the test using a stratified sample from New York.

The Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) was a revision of the Wechsler-Bellevue published because of the need for a more robust intelligence test for recruits during the World War II and a preference for the two score approach over the traditional one score approach used by Binet (1905). A large standardisation sample across the age range (ages 16-75) was used. It was then replaced by the Wechsler Adult Intelligence Scale – Revised (WAIS –R; Wechsler, 1981) which was similar to its predecessors in supporting a verbal – performance distinction. Like the previous two Wechsler scales, the WAIS-R had six subtests making up the Verbal Scale and five subtests making up the Performance Scale; the eleven combining to give an overall intelligence score known as Full Scale IQ. However, evidence was gathering from factor analytic research that adopting a verbal and performance structure potentially ignored other dimensions of intelligence (Cohen, 1957). Analysis of WAIS-R subtest scores provided evidence of a third factor, often described as ‘freedom from distractibility’, which tapped into subtests which seemed to require a working memory component (Arithmetic, Digit Span and Digit Symbol). Further evidence suggested a fourth factor ‘processing speed’ should also be distinguished from verbal and performance scores particularly given the slower processing speeds of some groups such as the elderly. These findings suggested a new measure which incorporated these additional components would not only enhance the validity of the Verbal and Performance scores but also provide insight into other hypothesised components of intelligence (Arnau & Thompson, 2000).

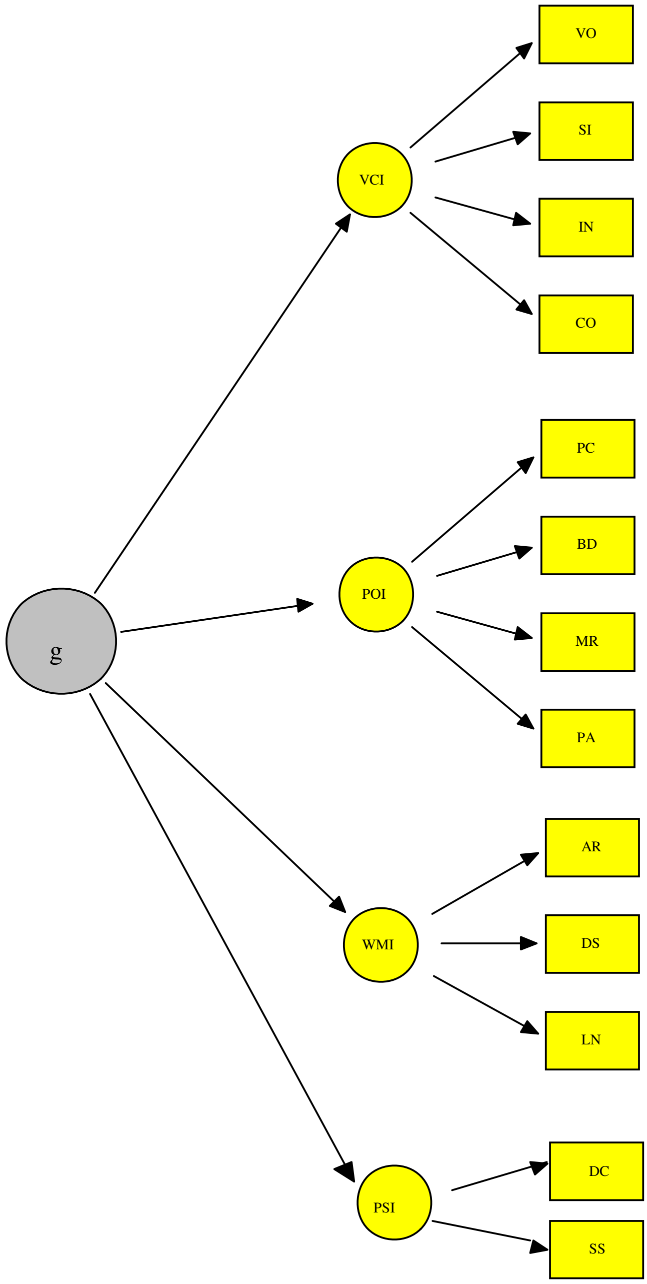
In 1997 a new version of the WAIS-R was published and was referred to as the Wechsler Adult Intelligence Scale – 3rd Edition (WAIS-III; Psychological Corporation, 1997). The main goals of the WAIS-III were retaining the basic structure of the Wechsler scales and retaining the use of deviation IQs, also known as standard scores. These were preferred to Terman's original computation of IQ. Eleven original Wechsler-Bellevue subtests were retained and three additional subtests were included. The standardisation sample was 2450 American individuals matched in age, gender, geographic region and educational level to the 1995 United States Bureau of Census data. Age related scaled scores on the WAIS-III allow comparison of scores to the standardization sample. Scaled scores have a mean of ten and a standard deviation of three and can be summed to give Full Scale IQ, Verbal IQ and Performance IQ (Psychological Corporation, 1997).

As a new addition to the Wechsler scales, four index scores were developed: Verbal Comprehension (VCI), Perceptual Organisation (POI), Working Memory (WMI) and Processing Speed (PSI), to reflect current theoretical developments in understanding, in particular, theory on working memory (e.g. Baddeley, 1986) and to have foundations in the results of factor analytic studies rather than clinical experience (Kaufman, 2000). The three IQ scores and the four index scores have a mean of one hundred and a standard deviation of fifteen and are calculated from the scaled scores of the individual subtests. The organisation of the four indexes of the WAIS-III according to The Psychological Corporation can be seen in Figure 2. [3] The test was adapted to counteract floor and ceiling effects and measure IQ scores of between 45 to 155. Extensive data on reliability and validity of the scale can be found in Ryan and Lopez, (2001). A summary of the WAIS-III subtests is provided in Appendix 1.

The WAIS-III aimed to move its foundations away from Wechsler's initial clinical perspective and base its construction more solidly in theory, while retaining its psychometric qualities (Psychological Corporation, 1997). Flanagan and Kaufman (2004) outline a framework described by Kamphaus and colleagues (1997) to describe the changes in how intelligence tests in general and the Wechsler tests in particular have been interpreted over time. The changes in interpretation can be described as four waves: 1) quantification of general level; 2) clinical profile analysis; 3) psychometric profile analysis; and 4) application of theory to intelligence test interpretation. Table 1 illustrates this evolution in interpretation.

[3] A similar construct of a verbal, performance distinction and four indexes is replicated in the Wechsler Intelligence Scale for Children – 3rd Edition (WISC-III). The Wechsler Intelligence Scale for Children (WISC) has a similar history to the WAIS in that its beginnings were in the form of Wechsler-Bellevue (Form II) which was extended to create the WISC (Wechsler, 1949) for the 5-15 age range. This was replaced by the revised version the Wechsler Intelligence Scale for Children (WISC-R; Wechsler, 1974), the Wechsler Intelligence Scale for Children – 3rd Edition (WISC-III; Wechsler,1991) and the Wechsler Intelligence Scale for Children – 4th edition (WISC-IV; Wechsler, 2003).

Figure 1. The organisation of the WAIS-III, showing second order *g*, the 13 subtests and their proposed relationship to the four indexes



The WAIS-IV (Psychological Corporation, 2008) is markedly different from previous Wechsler scales. It retains the four indexes (VCI, Perceptual Reasoning Index - PRI renamed from POI to match the Index on WISC-IV; WMI and PSI). The WAIS-IV no longer offers a Verbal and Performance IQ and the FSIQ is now calculated as the sum of four scales (three VCI, three PRI, two WMI and two PSI). Only eight of the original eleven subtests which made up the WAIS-III FSIQ remain. In addition, a global score has been added; reflecting the sum of the scaled scores on three VCI subtests and three PRI subtests. This change fits with the fourth wave of test interpretation and is said to reflect the latest theoretical advances particularly from the literature in intelligence theory, cognitive neuroscience and adult cognitive development (Flanagan & Kaufman, 2004). It focuses in particular on three constructs: *fluid reasoning*, the "ability to process or manipulate abstractions, rules, generalisations, and logical relationships", *working memory*, the " ability to actively maintain information in conscious awareness, perform some operation or manipulation with it, and produce a result" and *processing speed*, the " ability to process information rapidly (which is dynamically related to one's ability to perform higher-order cognitive tasks)"(Lichtenberger & Kaufman, 2009, pp.20-21).

Table 1: Changes over time in how intelligence tests are interpreted (modified from Flanagan & Kaufman (2004)

Wave	Key Characteristics
Wave 1: Quantification of general levels	Driven by need for classifying individuals into groups; focus on Global IQs; Wechsler talks of deviation from the mean; one score consistent with notion of *g*.
Wave 2: Clinical Profile Analysis	Interpretation of an individual's cognitive profile through performance on subtests; the Verbal- Performance dichotomy; apparent shifting by Wechsler between a global concept and separate mental abilities; interpretation of profiles to influence diagnosis and treatment but methodologically weak; less focus on global IQ.
Wave 3: Psychometric Profile Analysis	Factor analysis by Cohen (1959) led to interpretation based on 3 indexes on the WISC-R and WISC-III: Verbal Comprehension; Perceptual Organisation and Freedom from Distractibility; Four indexes introduced for WAIS-III; move away from individual subtest performance; greater psychometric expertise required in test interpretation; move towards a more solid evidence base but tests still criticised for lack of foundation in intelligence theory and empirical support.
Wave 4: Application of theory	Reorganisation of subtests into clusters defined by theory; WISC-III introduced new index: Processing Speed, but still criticised for poor link to theory; Introduction of cross-battery approach aimed at interpreting scales based on CHC theory; development of WISC-IV and WAIS-IV which claim to be more embedded in current theory.

Lichtenberger and Kaufman (2009) claim the popularity of the Wechsler tests remains 'remarkable and pervasive' (p.19). Rabin *et al.* (2005) found the WAIS adult scales the preferred test for measuring intelligence by clinical neuropsychologists, in assessment by forensic psychologists (Archer *et al.*,2006), clinical psychologists (Camara *et al.*,2000) and other psychologists (Groth-Marnat, 2009; Kaufman & Lichtenberger, 2006). The most current adult version of the Wechsler scales claims to continue the tradition of rigorous standardisation procedures and structural integrity (Lichtenberger & Kaufman, 2006).

Summary

The Wechsler scales provide clinicians with a psychometrically sound means of understanding the intellectual abilities of both children and adults across the life span. The publication of the WAIS-IV in 2008 is testament to the continued success and demand for the scales in the field of applied psychology. However, debate continues about the theoretical basis of the scales and the internal validity of the scale across all populations; particularly for people with learning disability who rely heavily on their use for gaining a diagnosis. The following section will explore these issues further.

3. The theoretical basis of the Wechsler scales

As theory has become more influential in the development of modern intelligence tests, it has become possible to assess the fit of models such as the Wechsler scales to current theories of intelligence. For example, the theory most used as the basis for new intelligence tests is Cattell-Horn-Carroll (CHC) theory. (Flanagan & Kaufman, 2004 provide a summary of how major intelligence tests map onto CHC abilities). However, questions remain about the structural validity of the Wechsler tests. A confirmatory factor analysis to assess the fit of their model to the standardisation sample carried out by the test publishers and reported in The Technical and Interpretative Manual of the WISC-IV (Psychological Corporation, 2003) did not specify factor loadings and correlations nor reported on information about the stability of the factor structure across the age range (Flanagan & Kaufman, 2004). Subsequent analyses (Keith *et al.,* 2006) found, when the standardisation data was analysed, better support for CHC theory and five factors than the four factor model proposed by The Psychological Corporation. In a bid to unite the two findings, Flanagan and Kaufman (2004) suggest using "Planned Clinical Comparisons"; eight new clinical clusters incorporating the findings of both Keith *et al.* (2006) and The Psychological Corporation. This would give clinicians information additional to the four indices, as well as offer an alternative model for interpretation (Flanagan & Kaufman, 2004).

Similar to the application of CHC theory, *gf-gc* theory has also been used to explain performance on the WAIS-III in terms of theoretical concepts of intelligence. Ryan *et al.* (2000) report age effects on WAIS-III subtests. They found little change with verbal subtests over time compared to a progressive decline in performance subtests (Kaufman & Lichtenberger, 1999). Scaled scores for most verbal subtests remained consistent across older and younger adults, however, scaled scores for performance subtests in younger adults were not maintained in older adults. Ryan *et al.* (2000) suggest this fits with the constructs of crystallized and fluid intelligence; the former showing more resistance to decline with age. Ardila (2007) found in all subtests, except digit span, an increase in score dispersions. It is suggested this points to increased heterogeneity in intellectual abilities with age, with greatest variability for subtests requiring executive functions, attention and some non-verbal abilities. Decline with normal ageing is thought to be more homogeneous for visuo-constructive abilities and general knowledge (Ardilla, 2007).

Kaufman (2000) suggests the limitations of the Wechsler scale become apparent when reference is made to some contemporary theories of intelligence, including

Sternberg's triarchic theory (1985, 1988, 1997) and Gardner's multiple intelligences theory (1983). The original clinical basis for their construction as opposed to a theoretical one does not, however, negate their usefulness and Kaufman (2000) argues that they have a place until such a time that: "there is something of value to replace them" (p.472).

3.1 Factor analysis and the Wechsler scales

Factor analytic studies have provided much of both the support and criticism for the Wechsler scales and their acclaimed validity. However the technique of factor analysis is not without its critics. Kline (1994) outlines some of these arguments but argues in favour of the approach and suggests if a wide range of variables are sampled a factor analysis can extract factors which had not been previously considered. Ford *et al.* (1986) carried out a systematic review of the technique in applied psychology and found that many of the decisions made by researchers about how to carry out such analyses were poor. In an attempt to provide guidance to researchers, Costello and Osborne (2005) offer a best practice guide for carrying out exploratory factor analysis. They emphasise the exploratory nature of EFA and suggest that many researchers wrongly use this technique to draw conclusions about the data when they should be using CFA; a process which allows researchers to capitalise on inferential statistics in testing hypotheses. CFA also has its weaknesses and these are outlined in the Method section.

For all versions of the WAIS, large standardisation samples were factor analysed. These samples are assumed to be normally distributed and representative of the general population. In comparison, clinical, abnormal or unusual populations are more likely to have reduced variance and thus any observed differences could be attributed either to their greater homogeneity or due to actual differences in how the scales are tackled by the group (Smith, 2003). These smaller studies with restricted groups of individuals can highlight subtle differences in performance which may otherwise be masked by a larger more heterogeneous sample (Burton *et al.*, 2002).

A literature review carried out by Smith (2003) identified 36 studies using exploratory factor analysis with the WAIS, WAIS-R or the WAIS-III between 1943 and February 2003. Twenty five factor analytic studies of the WAIS, WAIS-R and WAIS-III with clinical (e.g. schizophrenic); or unusual or abnormal populations (e.g. older adults or people with learning disability) were highlighted (Smith, 2003). Despite the differing methodologies and subjective interpretation common to factor analysis (Smith, 2003) some key findings emerged. The following section will look at these and other findings in more detail.

3.2 Factor analysis and the WAIS and WAIS-R

Hill *et al*. (1985) reviewed factor analytic studies of the WAIS and WAIS-R. Consistent with the Wechsler's proposed structure of the WAIS, they suggested a two factor verbal – performance solution was most reliable and there was little real evidence for a third solution. A review by Leckliter *et al*. (1986) grouped together studies of different clinical populations using the WAIS-R. Some researchers found only two factors could be reliably extracted from the data and that little additional variance was accounted for by a third factor (Atkinson & Cyr, 1984; Piedmont *et al.,* 1992).The verbal and performance dichotomy was also found to be the most reliable by Enns and Reddon (1998). A three factor model did not hold across the age groups and the researchers suggested the merit of using the third factor only in select clinical or research situations; a view supported by researchers such as Burgess *et al*.(1992); Ryan *et al.* (1997) and Clausen and Scott (1993).

Crawford *et al*. (1989) examined the factor structure of the WAIS-R in a UK sample in comparison to a standardisation sample from the UK. They found a similar structure in both populations but suggested the equal weightings given to the Verbal and Performance IQ scores should be questioned. In particular, the construct validity of the Performance IQ score appeared weak and led the authors to suggest a move towards a factor based scoring system which would lend itself more to analysis of neurological research samples. Sturmey *et al*. (1993), recognising the lack of clarity in the findings from clinical populations, also suggested interpreting results based on the factor analysis of specific groups rather than relying on normative data which may suggest a differing structure.

3.3 Factor analysis and the WAIS-III

The four index scores on the WAIS-III are intended to help provide a more detailed and clinically useful picture of an individual's strengths and weaknesses and can be considered the explicit theoretical structure of the WAIS-III. However, the scale also promotes an implicit factor structure which can be adopted for scoring purposes in line with FSIQ, VIQ and PIQ. The review by Smith (2003) suggested evidence for the WAIS-III explicit four factors is mixed. Kaufman *et al*. (2000) carried out factor analysis with the standardisation sample used in the WAIS-III. They specified two and three factor solutions *a priori* and found support for a verbal and performance solution as well as a three factor solution which included a factor they describe as executive functioning. They further suggested that this factor may be less valid in populations where components of this executive functioning factor may be discrepant, for example, when there are differences between working memory and processing speed. Despite these results, Kaufman *et al.* (2001) chose to align themselves with a four factor solution of the WAIS-III (Smith, 2003).

Support for the construct validity of the WAIS-III was also claimed in a larger Canadian sample (n= 718) by Saklofske *et al.* (2000). They suggest that their results from both exploratory and confirmatory factor analyses replicate the four factor solution and: "attest to the psychometric integrity of the WAIS-III and its "'portability' across cultural boundaries." (p.438). Garcia *et al.* (2003) looked at the factor structure of the WAIS-III in a Spanish sample (n=1369) using confirmatory factor analysis to check fit with a one factor, two, three and four factor and second order model. Fit was best for a four factor solution and support was also good for a second order factor consistent with *g*. Support was also found in Chinese mainland non-clinical and clinical samples (Yao *et al.* 2007). They highlight the significance of finding additional empirical support for the processing speed index, which provides a clinically meaningful measure distinct from the traditional WAIS-R three factor approach and appears particularly sensitive to brain pathology, as suggested by the test publishers. Ryan and Paulo (2001) found support for four factors from a small sample of patients referred for intellectual or neuropsychological assessment however suggested further work be carried out to confirm this finding in other diverse patient groups (Smith, 2003). Dickinson *et al*. (2002) also found support for four factors in a sample of people with schizophrenia; although they suggested the impairment caused by schizophrenia may have caused the individual to tackle certain subtests differently and resulted in a change in the pattern of loadings.

Taub (2001) investigated this explicit and implicit construct validity of the scale using CFA with a sample from the standardized data. The results supported a four factor and general factor model, however, did not support a verbal-performance dichotomy. The performance factor was found to be subsumed by a second-order *g*. It was suggested that practitioners should use a four factor model as the basis of their interpretation of first order factors and exercise caution in interpreting VIQ and PIQ differences (Taub, 2001). Support for a four factor model was also found in a confirmatory factor analysis of the standardization sample (Arnau & Thompson, 2000). This research claimed to be the first second order factor analysis of the WAIS-III acknowledging a second order *g* factor in addition to broader first order factors (the WAIS-III technical manual reports a first order CFA which favoured a four factor model, over a one, two and three factor model; Psychological Corporation, 1997). In Arnau and Thompson (2000) their second order CFA found support for a second order factor of *g* and four first order factors, leading the researchers to suggest that practitioners should use the index based factor scores rather than focusing on the traditional verbal and performance IQ scores. They suggest this would enable individual differences to be more sensitively measured.

This enhanced differentiating power of the test is assumed to operate across the population. However Burton *et al.* (2002) exercises a little more caution, suggesting:

"the necessity of examining patterns of WAIS-III latent variability within divergent clinical samples in order to derive hypotheses about the structure of intelligence that are specific to the characteristics of those samples. This becomes particularly necessary when clinicians use normative samples to calculate summary indices that imply a distinct pattern of latent variability among the WAIS-III subtests and that are then used as a basis to differentiate clinical disorders." (p.374).

Burton and colleagues' (2002) study of a confirmatory factor analysis of the WAIS-III in a clinical sample, cross validated with the standardization sample aimed to show which of nine hypothesised models best fit the data and to what extent the findings in the standardisation sample applied to their distinct clinical group. Burton *et al.* (2002) suggest the tests publishers should have evaluated the fit of more complex models before promoting the four factor model as the best fit. The nine hypothesised models (including the five original less complex models tested for fit by the test publishers) were assessed on goodness of fit in a clinical sample (n=328) and then cross-validated with the WAIS-III standardization sample. Results suggested a six factor model, corresponding to an expanded version of *gf-gc* theory provided the best fit to the data and was more accurate than the test publisher's model in explaining the latent variability among subtests. The six factors represented the following dimensions: Semantic Memory, Verbal Reasoning, Visual Reasoning, Working Memory, Constructional Praxis and Processing Speed.

Despite this conflicting finding, Burton *et al.* (2002) suggest this does not negate the usefulness of the four factor model. They claim there is enhanced clinical utility in interpreting these results if a hierarchical view of intelligence is considered and takes account of broad and narrow abilities. In this way, the interpretation of the WAIS-III will be most relevant to the individual being tested. For example, if an individual's scores on the 13 subtests do not differ substantially from the mean, it is appropriate to use FSIQ as a measure of general intelligence. Similarly, when scores for the subtests relating to the verbal IQ and the performance IQ do not differ greatly from their respective mean score, it would be appropriate to interpret a VIQ and PIQ score as a reflection of verbal abilities and performance abilities. This would also apply to the four indexes when there is limited variability across subtests. However, when there is increased 'intersubtest scatter', the individual IQs or indexes may not reflect a unitary construct and therefore the construct becomes uninterpretable. While there is no definitive guidance on at what point scatter of scores among subtests is too marked to be interpreted, some suggest this should be the equivalent of when scatter within the composite is equal to or more than for 95 per cent of the standardisation sample (Ryan& Lopez, 2000) or put another way: when the point difference between the scaled scores is very unlikely to have occurred in the standardisation sample (< 5 per cent). Burton *et al.* (2002) suggest when scatter to this degree occurs, there may be more clinical utility in interpreting scores based on their six factor model.

3.4 Factor analysis of the Wechsler scales with people with low IQ

Few factor analytic studies have been carried out with the Wechsler scales and people with learning disability; though more may have used the Wechsler scales for children (Smith, 2003). Sprague and Quay (1966) looked at scores on the WAIS and found three factors best fitted the data: verbal, freedom from distractibility and performance. Their analysis of 124 people with full scale IQ of less than 80 used a low criterion for factor loadings. Smith (2003) suggests that if a more robust criterion applied, no performance factor would emerge. Atkinson and Cyr (1988) also analysed a sample of individuals with a full scale IQ less than 80. Their study of the WAIS-R offered support for two and three factor solutions. The results supported the verbal and performance distinction and an additional freedom from distractibility factor. Ryan *et al.* (1992) also found support for the verbal and performance distinction but suggested that the arithmetic subtest was 'factorially complex". The review found no study examining the factor structure of the WAIS-III in a population of people with a learning disability.

3.5 Construct validity of the WAIS-III for a low IQ population

In order to investigate the relevance of the factor structure of the standardisation sample to a low IQ population, Jones *et al.* (2006) carried out an exploratory factor analysis of 105 WAIS-III scores from a population with low IQ . The data was collected from routine clinical practice with 105 individuals with a full-scale IQ of 74 or below. The factor analysis revealed only one robust solution, representing verbal and performance factors. They concluded the four indexes, extracted from standardization samples and which have traditionally been used to inform practitioners about the profile of individuals, were not supported by their research on a population with low IQ.

The sample analysed by Jones *et al.*, (2006) was small (n=105) however it did meet a number of criteria: number of cases $\geq$ 5x number of variables (Lewis, 1995), N-n-1 $\geq$

50 (where N = number of participants and n = number of variables; Lawley & Maxwell, 1971), and a ratio of at least 2:1 participants to variables (Kline, 1994).

Several preliminary tests were carried out to assess suitability of the data for factor analysis and given the lack of normality and linearity the findings were not felt to be unusual for a clinical sample (Jones, *et al.* 2006). Seventy percent of the correlations present in the matrix were at least 0.3. Sampling adequacy was found to be high at 0.861. Based on the recommended size of correlations (Tabachnick & Fidell, 2001) the data set was considered suitable for factor analysis.

The 'verbal' factor accounted for the following subtests: vocabulary, similarities, comprehension, digit span, information, arithmetic, letter-number sequencing. The 'performance' factor accounted for: digit-symbol coding, picture arrangement, symbol search, picture completion and matrix reasoning subtests. The analysis revealed a lack of simple structure for the three-factor and four factor solutions. However, the two factor solution showed a simple structure and identical pattern of loadings greater than 0.45. Between factor correlations were 0.642 and 0.532 and this accounted for 40 per cent to 48 per cent of the variance.

The research concluded a two factor solution - verbal and performance was the only reliable solution suggesting the WAIS-III potentially had reduced neuropsychological utility for individuals with low IQ. Jones *et al.* (2006) suggested index scores on the WAIS-III should be used with caution though the use of the verbal and performance scores were justified. The study used individuals with low

IQ (IQ ≤ 74) which suggests the results could be generalised to individuals with borderline intellectual functioning (IQ 70-79) and also people with learning disability. The study suggested a need for a further confirmatory work on the construct validity of the four factors for a larger group of people with low IQ.

Summary

Many factor analytic studies of the Wechsler scales have been carried out with both the standardisation samples and clinical samples. Some results have supported the test publisher's findings on the construct validity of the scales while others have not. Similarly, support for the validity of the four indices of the WAIS-III has been found in some clinical populations (Heijden & Donders, 2003; Ryan & Paulo, 2001; Dickinson *et al.,* 2002) while other research (Burton *et al.*, 2002) suggests a more complex six factor model provides a better fit for some populations. As far as is known, there has only been one study looking at the factor solution of the WAIS-III in a sample of people with low IQ (Jones *et al.* 2006) and this has cast doubt on the validity of four indices in this population.

4. Using the WAIS-III with individuals with a learning disability

In addition to concerns about the internal validity of the WAIS-III there are other concerns regarding the use of the scale with people with a learning disability. These will be outlined in this section.

4.1 Administering the WAIS-III to people with a learning disability

The Wechsler scales have been criticised for their lack of appropriateness for individuals with a learning disability (Whitaker, 2005). Testing can be anxiety provoking for many individuals and for those with a shortened attention span the long administration time can pose a significant challenge. To address these issues, the WAIS-R was abbreviated to dyad, triad and tetrad tests, each showing good reliability and validity and all averaging less than 20 minutes administration time (Kaufman, Ishikuma & Kaufman-Packer, 1991). Short forms of the WAIS-III have also been developed. In an evaluation of the utlility of three abbreviated forms in clinical populations: the Satz-Mogel abbreviation (Satz &Mogel, 1962); a seven subtest short form (Ward, 1990) and a clinically derived abbreviation, all forms were highly correlated with the full form. The closest approximation to full scales scores for the tested population was found when using the clinically derived abbreviation (Wymer *et al.* 2003). Despite the lack of standardisation of some short forms, they are often preferred by clinicians, who are comfortable with the subtests and may not have the budget to buy new tests (McKenzie & Paxton, 2006).

The Wechsler Adult Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) claims to be a standardised and validated short form of the WAIS-III (WASI manual, 1999). Uses for the WASI include screening, estimating cognitive functioning and developing clinical interventions (Clayton *et al.* 1986). It has an advantage over other short forms by having its own independent norms. However, concerns have been raised and acknowledged in the manual (WASI, 1999) about its accuracy when used to assess intellectual ability with people with a learning disability (Clayton *et al.*

1986; Crawford *et al.* 1992). Although people with a learning disability (individuals with full scale, verbal and performance IQs all under 70) were used in the standardisation sample of the WASI, the numbers were small and level of functioning was not specified. Thirteen percent of the group when tested using the WASI did not fall below the cut off of IQ 70 suggesting the WASI would misdiagnose some individuals (McKenzie & Paxton, 2006). Abbreviated forms will continue to be used in populations where full administration of the most current WAIS scale is neither practical nor sensitive to the needs of the individual, as is often the case with individuals with a learning disability. In such scenarios, practitioners may need to be wary of using short forms to formally classify intellectual functioning (McKenzie & Paxton, 2006). Similarly, screening tools for learning disability are now available and while not intended to replace full diagnostic assessments of learning disability they may provide a quicker route into apportioning of initial services to those who need without the need to carry out lengthy assessments (McKenzie & Paxton, 2006).

Despite the psychometric credentials of the WAIS-III, research suggests a high number of errors are made by those trained to use the tool. In a survey of clinical psychologists working in the field of learning disability in Scotland, McKenzie *et al*. (2004) found 83 per cent did not follow the standardised instructions and more than 75 per cent did not use all subtests; a practice known to increase the standard error of measurement. Despite the finding that non standard administration can increase IQ scores (Joncas & Standing, 1988; Slate *et al.*, 1991), for some populations it has become a clinically appropriate means of testing (Groth Marnat *et al.*, 2000). Some of these modifications are outlined by Hishinuma (1998).

4.2 Standardisation of the WAIS-III in a sample of people with learning disability

One problem with interpretation of the WAIS III with people with a learning disability relates to the underrepresentation of people with a learning disability in the standardisation sample. Whitaker (2005) comments on the large stratified samples used for standardisation of the WAIS-III. In the US sample, 200 adults per age group were used and this would coincide with only five people scoring two standard deviations below the mean. Moving further from the mean score to 2.67 standard deviations below the mean, would result in only one person per age group in the sample. The typical profile of abilities found in the mild learning disability standardisation sample has also been found to differ from some clinical populations. In a clinical sample, Murray *et al.* (2003) found a higher proportion of individuals with better verbal skills; with some individuals gaining scores on VIQ above 70. This problem of representativeness of the standardisation sample is further enhanced by the exclusion criteria applied to some individuals, such as those with a sensory impairment. The reliability and validity of this form of testing for people with very low intellectual functioning is also questioned by the British Psychological Society (BPS, 2001). Wechsler himself made it clear the original tests were first and foremost clinical rather than psychometric tests made to assess those with average or near-average intelligence and not those with IQs below 70 or above 130 (Lichtenberger & Kaufman, 2009).

Whitaker (2005) also highlights some of the concerns of the accuracy of the WAIS-III for people with low IQ suggesting the conversion of raw scores to scaled scores does not accurately differentiate those in the bottom 0.13 per cent of raw scores. For example, it is possible to gain a raw score of zero by failing to gain any score on a subtest. However, the equivalent scaled score is given as one. This means an individual's score may be an overestimate of their ability. Whitaker (2008) equates this with an individual whose ability is more than three standard deviations below the mean is rated as if their ability is three standard deviations below the mean and suggests caution should be exercised when interpreting IQs or index scores based on scaled scores of one. Despite these limitations, the WAIS-III continues to be used extensively with people with low IQ and for many people remains part of the key to accessing critical support and care.

4.3 The Flynn Effect

The Flynn effect refers to the finding that over the twentieth century there have been large gains in IQ from one generation to the next (Flynn, 2009). Evidence from the WISC demonstrates how some subtests (irrespective of their assumed relative loading on *g)* have shown larger gains over time than others. Flynn (2009) suggests children today are more adept at using abstractions and logic to solve novel problems. Our ancestors used more concrete rules to govern their problem solving. As such, some subtests have seen massive gains because societal trends have favoured the skills required to complete them. This finding has a detrimental effect on the value of the published normative data and in order for gains to be accounted for new norms are introduced every 15-20 years. Differences consistent with the Flynn effect have been demonstrated in the WAIS-IV. The average WAIS-IV FSIQ score was found to be 2.9 points lower that WAIS-III FSIQ. The differences in index scores were found to be greater for the Verbal Comprehension Index and the

Perceptual Organisation/Reasoning Index with index scores for the WAIS-IV lower than for the WAIS-III. These lower scores are considered to be a more accurate representation of an individual's ability because the norms are more current (Lichtenberger & Kaufman, 2009).

The Flynn effect has also been found in other studies. In a population of school children with low IQ, Kanaya *et al.* (2003) found the likelihood of a diagnosis of learning disability was influenced depending on the test and norms used. In the early years of new norms, the test was found to be harder and as a result more individuals fell into the IQ sub 70 range. The researchers found nearly 20 per cent of a group of children with borderline IQ were classified as having learning disability on retesting with the WISC-III compared to the previous test using the WISC-R. Fitzgerald *et al.* (2007) found full scale IQ scores were typically four or more IQ points lower on the WAIS-III than on the WAIS-R. This would lead to 66 per cent more people meeting the criteria if they had been assessed using the newer and more difficult WAIS-III scale. Assuming a UK population of six million, the authors suggest this would result in an extra 480,000 people in the UK diagnosed with a learning disability. In situations where IQ test scores determine access to services, the year of test administration could be crucial. However in the case of convicted criminals in the United States, year of administration could be the difference between a life sentence and the death penalty (Kanaya *et al.* 2003).

"The main conclusion that can be drawn from these results is that caution should be used when basing important financial, social or legal decision on IQ scores," …. "Perhaps the most important times to be particularly cautious are when a test is either at the beginning or at the tail end of its norming cycle. Although test scores are most valid at the beginning of a norming cycle, they run the greatest risk of being compared to highly inflated scores from the waning years of the previous norming cycle." (Kanaya *et al.*, 2003. p.789).

4.4 Diagnostic cut-off

Greenspan (1999) discusses the changes in IQ cut-off for learning disability services. In 1973 the cut-off in the United States changed from one standard deviation below the mean (eighty-five) to two standard deviations below the mean. In 1992 the American Association for Mental Retardation increased the cut-off point from seventy to seventy-five. These changes have an important knock-on effect on the cost and provision of services and, as Greenspan (1999) argues, an arbitrary cut-off influences our perception of people with a learning disability; promoting the notion that it is a homogeneous group with similar needs; potentially failing those who have real need but do not meet the cut-off. The re-norming exercises of the Wechsler scales suggest the cut-off is largely arbitrary by ensuring the continued normal distribution of scores across the population. Thus, two per cent of the population will fall below a FSIQ of 70, and 50% of the population will fall under a FSIQ of 100 (Wechsler, 1997). However, Whitaker (2008) also highlights a discrepancy with the percentile ranking of the scale suggesting there are more people with severe and profound learning disability (0.4%) than suggested by a theoretical normal distribution (<0.1%). This underestimation suggests a learning disability prevalence rate of 2.28% for IQ <70. In reality, the true rate may be closer to 2.58%.

Psychologists are often involved in classifying learning disability given their general training in psychometric testing and the Wechsler scales in particular. Standardised assessment has provided the means to quantify intellectual impairment and allow services to adopt a more defined gate-keeping role should they wish. The precise cut-off above which services can be witheld can reduce the demand on already overstretched budgets or personnel. In practice, this may not always be the case. Psychologists may use their discretion to see individuals who are above the threshold but whom they know will be better served within a learning disability service. The BPS (2002) recommends psychologists use their informed clinical judgement around cut-off points and as such the cut off IQ below 70 should serve more as a guide than a fixed point above which individuals will automatically be denied certain services. However, adopting a fixed cut-off point could have major implications for individuals, for example, access to specialist day services, financial support, protection in the form of statutory legislation and even an individual's right to be a parent (Aunos *et al.* 2005).

Summary

As outlined in a previous section, there are alternative theories of intelligence which favour a more ecologically valid approach to intelligence testing. However, these have not been used extensively in determining intellectual functioning for diagnostic purposes. The role of the Wechsler scales in determining whether or not a person has a learning disability remains crucial. Despite this, its appropriateness for this client group is questioned by several authors and not least by Wechsler himself (Lichtenberger & Kaufman, 2009). The administration time of the test and complexity of some of the verbal instructions are not suitable for many individuals, particularly those with a learning disability. Alternative shorter forms may be more appropriate however may not provide the accuracy necessary to ensure appropriate classification of intellectual functioning (McKenzie & Paxton, 2006). The representativeness of the standardisation sample for this group has also been questioned (Whitaker, 2005; Murray *et al*. 2003).

Diagnosis can be hugely important to individuals in terms of the support they may receive and the potential stigma attached to learning disability (DoH, 2001) Therefore the importance of getting the correct information to inform a potential diagnosis is critical, a process potentially complicated by the Flynn effect.

Debate continues about the cut-off point for intellectual functioning and whether this is an arbitrary point on the scale of intellectual functioning or reflects actual difference in cognitive profile between those above and below the cut-off. An exploratory examination of the internal construct validity of the Wechsler scales in order to check the fit of the proposed internal structure of the WAIS-III for a low IQ population has been carried out by Jones *et al*. (2006) and suggests the four index structure has poor validity in this population. By employing confirmatory measures it might be shown whether this structure holds for people with borderline IQ and those with a learning disability.

5. Rationale and aims of the study

5.1 Rationale

A number of authors (Flynn, 1984; Greenspan, 1999) argue that using a cut-off point for a diagnosis of learning disability is essentially arbitrary and does not account for factors such as IQ drift (Flynn, 1984) or practitioners using out-dated tests (McKenzie & Paxton. 2006). Nor does it acknowledge that individuals with a learning disability may not have a uniform cognitive profile (Murray *et al.*, 2003, Kaufman & Lichtenberger, 1999). Nevertheless, a diagnosis is often required for individuals to receive essential funding and services. Some services have adapted the criteria for learning disability to try to account for IQ drift and increased the cut off score to 75 (The American Association of Mental Retardation, 2002) or even 80 (Evers & Hill, 1999).

A diagnosis is one end result of using the WAIS-III. However, the WAIS-III is often used clinically as a tool for describing the pattern of strengths and weaknesses of individuals in order to specifically tailor services to an individual. The results on the four indices can be interpreted to give a general pattern of functioning that could be made more precise with closer interpretation of performance on subtests. Some research supports the explicit construct validity of the four indices suggested by the test publishers, while other research favours more complex five factor models or six factor models consistent with updated *gf-gc* or CHC theory. However, previous exploratory research with people with a low IQ by Jones *et al.*,(2006) supports the implicit two factor solution over the explicit four factor solution. The sample used in the Jones *et al.* (2006) research straddles three groups of intellectual functioning above and below the cut-off point (IQ 74 or below). Thus the findings call into question the validity of scores on the WAIS-III for those with significant learning disability (IQ 55-69) significant impairment of adaptive functioning and onset before age 18), those with severe learning disability (IQ<55), significant or severe impairment of adaptive functioning and those whose scores fall in the borderline range (IQ = 70-79). The Jones *et al.* (2006) study suggests that future studies should try to replicate this finding with a larger population and employ confirmatory methods to allow testing of the theory generated by their exploratory factor analysis.

To the author's knowledge there is no research which has examined whether there is an underlying difference in the factor structure of the scores obtained on the WAIS-III for those with learning disability (in the severe and significant range) and those who fall within the borderline range of intellectual functioning. Should the cut-off (IQ less than 70) be arbitrary, an analysis of the factors in the WAIS-III in a population with borderline IQ (70-79) should be the same as those with IQ less than 70. In contrast, a difference in factor structure could suggest that the cut-off is reflective of an actual difference in the cognitive profile of these two groups, as measured by the WAIS-III. This would have implications for how learning disability is defined theoretically and clinically, and for service provision.

5.2 Aims

The aim of the study is to test the best fit factor solution for a group with borderline IQ and those with severe and significant learning disability. The resulting factor solutions can be compared to factor solutions in published data on standardised samples, other groups and from recent research with individuals with low IQ (Jones *et al.* 2006).

The research will add to the discussion about the intellectual profile of individuals with low IQ and whether the cut-off point used to define learning disability reflects an actual difference as measured by the WAIS-III. The implications of the findings for the new WAIS-IV will also be discussed.

6. Method

6.1 Ethical Issues

The study aimed to gather pre-existing data from WAIS-III scores collected routinely by Clinical Psychologists as part of their assessment process. The data was gathered retrospectively and no contact with patients was required. As such, no consent was gained from the individuals to use the data for research purposes. The ethics committees of the National Health Service and Edinburgh University granted permission for the work conditional on approval from the relevant Caldicott Guardian. This approval was granted (see Appendix 2). The study seeks to investigate potential differences in cognitive profile as measured by the WAIS-III between a sample of people with a learning disability and a sample of people with borderline IQ. The outcomes of the research have potential implications for the way the WAIS-III and future versions of the Wechsler scale is used with people with a learning disability, which in turn can impact on the provision of services. Thus it was considered important to remain sensitive to these issues when reporting these results. Clinicians were informed that the results of the study would be made known to them. They were also informed that the results would be written up in journal format and submitted for publication to raise professional awareness of some of the issues.

6.2 Recruitment of Psychologists

The Lead Psychologist for a learning disability service in Southern Scotland was emailed information regarding the study and asked if they wished to be involved. Once permission was given for data to be collected from the service the Lead Psychologist provided contact details for the four other Psychologists in the service who would be able to provide access to WAIS-III scores. Telephone and email contact was made with these individual Psychologists to brief them about the research and to arrange access to their files.

6.3 Procedure

Historical and current files stored in each learning disability service were reviewed by the researcher and information was gathered from those files where a WAIS-III had been scored. Information from unscored tests was not included as these would not have provided the required scaled scores.

The Data Recording Sheet (Appendix 3) had three sections:

1) Demographic details. Information was collected on gender and age.
2) WAIS-III subtest scores. The form included space for age-scaled subtest scores.
3) Identifying code. An identifying code recorded the initials of the clinician whose service the file had come from and a ‘data set number’. This was recorded on the data recording sheet and also on a Clinician’s Recording Sheet (Appendix 4). The Clinician’s Recording Sheet contained the name of the client next to the identifying code. This was to be kept by the clinician

until the end of the research and was not kept by the researcher. This document was to ensure data could be extracted again if lost during the inputting procedure and also to ensure entries were not duplicated.

6.4 Data Collection

The data was collected over a six month period. The WAIS-III was published in 1997, and data gathered was from the period 1997 -2010. At the end of the data collection period in each service, the psychologist was asked to keep the Clinician Recording Sheet confidential. They were told the researcher would contact them in due course to let them know when the sheet could be destroyed. They were given contact details of the researcher for any further questions and were told they would be informed of the outcome of the research.

6.5 Materials

Wechsler Adult Intelligence Scale – Third Edition (WAIS-III)

The WAIS-III contains 14 subtests, outlined in Appendix 1. Full details of administration and scoring procedures can be found in the WAIS-III manual (Psychological Corporation, 1997). The participant's raw scores are transformed into scaled scores and with reference to age related norms these are turned into IQ scores. The product of this scoring process should be a Full Scale IQ (FSIQ), a Verbal Scale IQ (VSIQ) and a Performance Scale IQ (PSIQ). In addition, scores for four indices can be calculated – Verbal Comprehension Index (VCI), Perceptual Organisation Index (POI), Processing Speed Index (PSI) and Working Memory Index (WMI). Table 2 shows the subtests used to calculate IQ scores and index scores. The Object Assembly subtest is the only subtest which does not contribute to the calculation of IQ or Index scores. Research, including the standardisation sample analyses, has therefore tended to exclude this subtest in factor analyses (Smith, 2003). In the current study the subtest is also excluded.

6.6 Reliability of the WAIS-III

The Psychological Corporation (1997, p.47) provides data on the reliability of the WAIS-III. The correlation coefficients for split half reliability, and using the Spearman-Brown formula, for each subtest, IQ and Index, range from 0.70 to 0.98. Test-retest revealed stability coefficients from 0.70 to 0.92 for the scores and inter-rater reliability show coefficients in the high 0.90s. These findings suggest the WAIS-III has good internal reliability as well as external reliability, i.e. an individual should perform similarly on two different occasions or if rated by different people. However, as stated in the introduction, it is acknowledged that administration of the WAIS-III to a learning disabled population will at times deviate from standardised practice (Groth Marnat *et al.*, 2000, Hishinuma, 1998).

Table 2. WAIS-III IQs, Indices and Subtests

Subtest	Index				IQ		
	VCI*	POI*	PSI*	WMI*	VIQ*	PIQ*	FSIQ*
Picture Completion		X				X	X
Vocabulary	X				X		X
Digit-Symbol Coding			X			X	X
Similarities	X				X		X
Block Design		X				X	X
Arithmetic				X	X		X
Matrix Reasoning		X				X	X
Digit Span				X	X		X
Information	X				X		X
Picture Arrangement						X	X
Comprehension					X		X
Symbol Search			X				
Letter Number Sequencing				X			
Number of Subtests	3	3	2	3	6	5	11

Note: X marks subtests that are included in calculations

- VCI – Verbal Comprehension Index, POI – Perceptual Organisation Index, PSI – Processing Speed Index, WMI – Working Memory Index, VIQ – Verbal IQ, PIQ – Performance IQ, FSIQ – Full Scale IQ (amended from Smith, 2003).

6.7 Validity of the WAIS-III

Data on the validity of the WAIS-III is also provided by The Psychological Corporation (1997). The tests claims to have good content validity in that it appears be consistent with neuropsychological concepts in the literature. Similar performance on WAIS-III and other scales claiming to measure intelligence would suggest that criterion related validity is good (coefficients from 0.68-0.93). Construct validity is the ability of a test to measure a theoretical construct or trait. All the items on the WAIS-III were found to show some association with each other; a finding consistent with the notion of *g*. The verbal and performance subtests correlated higher within their own scale than between them. This held for all subtests apart from arithmetic and picture arrangement which were found to have split loadings. However, Burton *et al.* (2002) suggest the analysis provided by The Psychological Corporation (1997), while finding best fit with the four index model, does not adequately assess the construct validity of the test by failing to assess more complex models of fit. Data from half the standardisation sample was used to assess the fit of five models including one factor general intelligence model; a two factor model representing the verbal and performance dichotomy; a three factor model (Verbal Comprehension, Perceptual Organisation and Attention; a four factor model (as previously outlined) and a five factor model which expanded the four factor model to include an additional factor described as Quantitative Ability. Jones *et al*. (2006) also question the validity of a four factor solution.

6.8 Participants

The data came from individuals who had at one point in time used the services of the learning disability teams. The sample can be regarded as opportunistic. The data came from individuals who had completed the WAIS-III and who had a Full Scale IQ of 79 or less. This value allowed data to be collected for three groups: 1) a group of people with severe intellectual impairment (IQ < 55), 2) a group of people with significant intellectual impairment (IQ 55-69) and 3) a group of people with borderline intellectual impairment (IQ between 70 and 79).

Certain assumptions were made about the groups of people with a learning disability. All information came from a learning disability service and met criteria of significant (IQ 55-69) or severe impairment (IQ < 55) of intellectual functioning however it was not possible to assess if the individual met the other two criteria for a diagnosis of learning disability (significant impairment of adaptive/social functioning and age of onset before adulthood; BPS, 2002). According to the clinicians in the service, the individual would only have been seen by the service if meeting the service criteria. While in some isolated cases this may have meant an individual without a learning disability was seen this was felt to be acceptable given, for the purposes of this study, a true clinical population was sought. This differs from the Jones *et al.* (2006) study which excluded participants who had an acquired brain injury, epilepsy or an autism spectrum disorder. Smith (2003) outlines the rationale for this. While the inclusion of these groups might confound to some degree the results of the factor analysis, it was felt that such exclusions would diminish the authenticity of a clinical sample.

6.9 Sample size

Not all of the WAIS-III assessments were completed in full. Tables 3,4 & 5 show that the majority of missing data came from the subtests symbol search and letter number sequencing and as such can be considered non random. As can be seen from Table 2 the index scores can still be calculated without reference to symbol search and letter-number sequencing. Thus it is a real possibility that administrators have deliberately chosen to leave out these two subtests and only report index scores. To confirm this view, missing data analysis (SPSS version 17.0) was carried out which confirmed the data was not missing at random. On this basis, the data with missing values was omitted from the data set and this resulted in a sample size of n= 140 for the sample of people with severe learning disability and a ratio of cases to observed variables of 11:1. The sample size for the people with significant learning disability was n= 264 giving a ratio of cases to observed variables of 20:1. The same process was carried out for the sample of people with borderline IQ, leaving a data set of n=89. The ratio of cases to observed variables is 7:1. All samples met the criteria number of cases > 5x number of variables (Lewis, 1995) and N-n-*1* $\geq$ *50 (where* N = *number of participants and* n = *number of variables;* Lawley & Maxwell, 1971).

Table 3. Missing data by subtest for sample of people with severe learning disability

Subtest	PC	VO	DC	SI	BD	AR	MR	DS	IN	PA	CO	SS	LN
Number missing	0	0	4	0	2	0	0	1	0	4	4	25	31

Table 4. Missing data by subtest for sample of people with significant learning disability

Subtest	PC	VO	DC	SI	BD	AR	MR	DS	IN	PA	CO	SS	LN
Number missing	0	0	0	1	1	0	0	0	0	4	4	37	47

Table 5. Missing data by subtest for sample of people with borderline IQ

Subtest	PC	VO	DC	SI	BD	AR	MR	DS	IN	PA	CO	SS	LN
Number missing	0	0	1	0	1	0	1	0	1	3	2	23	18

An opportunity sample of 493 data sets were analysed. Data was obtained from 76 males and 64 females in the sample of people with severe learning disability. There were 161 males and 149 females in the sample of people with significant learning disability. In the sample of people with borderline IQ there were 57 males and 32 females. The frequency of data per age band and group is shown in Table 6. The mean age for the sample of people with a severe learning disability was 38.5 (standard deviation = 14.9). For the sample of people with significant learning disability the mean age was 31.3 (standard deviation = 13.6). The mean age for the sample of people with borderline IQ was 28 (standard deviation = 12.3). The age bands used replicated those used in the WAIS-III scoring manual. Table 6 describes the frequency and percentage of sample in each IQ band.

Table 6. Frequency and percentage of data per age band

Age Band	Frequency in sample of people with severe learning disability (n=140)	Percentage in sample of people with severe learning disability	Frequency in sample of people with significant learning disability (n=264)	Percentage in sample of people with significant learning disability	Frequency in sample with borderline IQ (n=89)	Percentage in sample with borderline IQ
16-17	5	3.6	30	11.3	12	13.5
18-19	5	3.6	34	12.8	19	21.3
20-24	20	14.2	43	16.3	21	23.6
25-29	17	12.1	21	8.0	7	7.9
30-34	17	12.1	44	16.8	5	5.6
35-44	26	18.5	42	15.9	11	12.4
45-54	29	21.0	30	11.3	10	11.2
55-64	14	10.0	17	6.4	4	4.5
65-69	2	1.4	2	0.8	0	0
70-74	3	2.1	0	0	0	0
75-79	2	1.4	1	0.4	0	0
80-84	0	0	0	0	0	0
85-89	0	0	0	0	0	0

Table 7. Frequency and percentage of sample in each IQ band

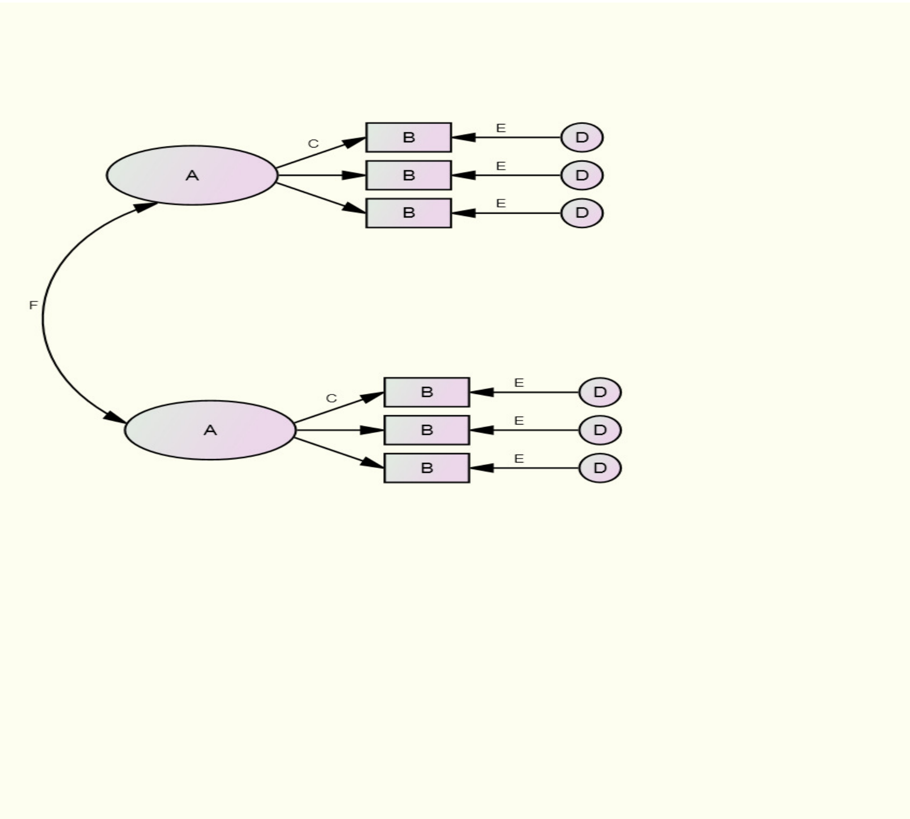
IQ level	Frequency	Percentage (%)
IQ 70-79 (Borderline range)	88	18
IQ 55- 69 (Significant LD range)	264	54
IQ < 55 (Severe LD range)	140	28

7. Analysis

7.1 Confirmatory factor analysis (CFA)

CFA is a process by which a data set can be used to check for theoretical fit against hypothesised models (Kline, 1994). CFA models can be drawn as path diagrams in which the observed variables are represented as rectangles and the latent variables (factors) as circles. Arrows with single heads represent the assumed direction of causality, whereas arrows with double heads suggest factors co-vary. Factors can point to more than one observed variable. Factor loadings express the effect (regression slope) of the factor on the observed variable while the squared factor loading or communality is the proportion of the variance of the observed variable that is explained by the latent variable. Unique factors contain the remaining variance (e.g. measurement error) once the variance caused by the latent factor is accounted for. These unique factors are linked only to individual observed variables. Figure 2 shows an example path diagram for a CFA.

Figure 2. Example path diagram



Key

A= latent variable, B= Observed Variable, C= Factor Loading, D = Unique Factor, E = Unique variance, F= Factor covariance

7.2 Sampling for factor analysis

There are a number of different views on the adequacy of the sample size required for factor analysis. Nunally (1978) suggests a ratio of 20:1 subjects to variables is required to avoid chance effects. Lewis (1995) suggests the number of cases $\geq$ 5x number of variables while others have suggested, $N-n-1 \geq 50$ (where N = number of participants and n = number of variables; Lawley & Maxwell, 1971 cited in Smith, 2003).

However, the general rule of thumb is the more the better (Costello & Osborne, 2005). Kline (1994) suggests that heterogeneous samples increase the variance and increase the loadings on factors. In an example, he explains how a factor analysis of abilities and the academic success of a homogeneous highly intelligent sample would not provide high loadings on intelligence because all individuals are intelligent and would experience academic success. Differentiation would be more apparent when considering factors such as interest and flair for subjects. A heterogeneous IQ sample on the other hand would show increased factor loadings on intelligence and would hold more face validity when looking across the whole population. Heterogeneity is generally good if the sample represents a real population and is not mixed just to increase sample size. However, there are times when differences are suspected between specific groups. In these instances determining the factor structure of a homogeneous group is important. (Kline, 1994).

Costello and Osborne (2005) found only 10% of small samples (where ratio of subjects to variables is 2:1) yielded correct solutions. In comparison, 70% of samples with a subject to variable ratio of 20:1 yielded correct factor solutions. In the study by Jones *et al.* (2006) subject to variable ratio was 8:1 (n=105) which when considered against Costello and Osborne's figures would yield a correct factor solution in approximately 50% of samples. However, they also recognise other conditions which can compensate for sample size; an argument further pursued in McCallum *et al.* (1999) who suggest that instead of focussing on sample size or ratio of sample size to variables, the level of communality ("the portion of the variance of that variable accounted for by the common factors"; p.85) of the variables and the number of indicators per factor are better indicators of required sample size.

When specifically examining the method of CFA, (Tabachnick & Fidell, 2001) suggests there are a number of weaknesses (studied in more detail in the Discussion section). A good fit between the data and the target matrix does not exclude the data from fitting other target matrices. The solution found is only one possible fit for the data and not necessarily the only one. In addition, the reliability of the statistical tests for rejecting or accepting the hypothesis can be questioned. There are a number of goodness of fit tests and like most statistical tools they all have their weaknesses. Kline (1994) also suggests a target matrix can be difficult to establish if the theoretical background is weak. However, this is largely circumvented if the target is a factor analysis of another sample already carried out.

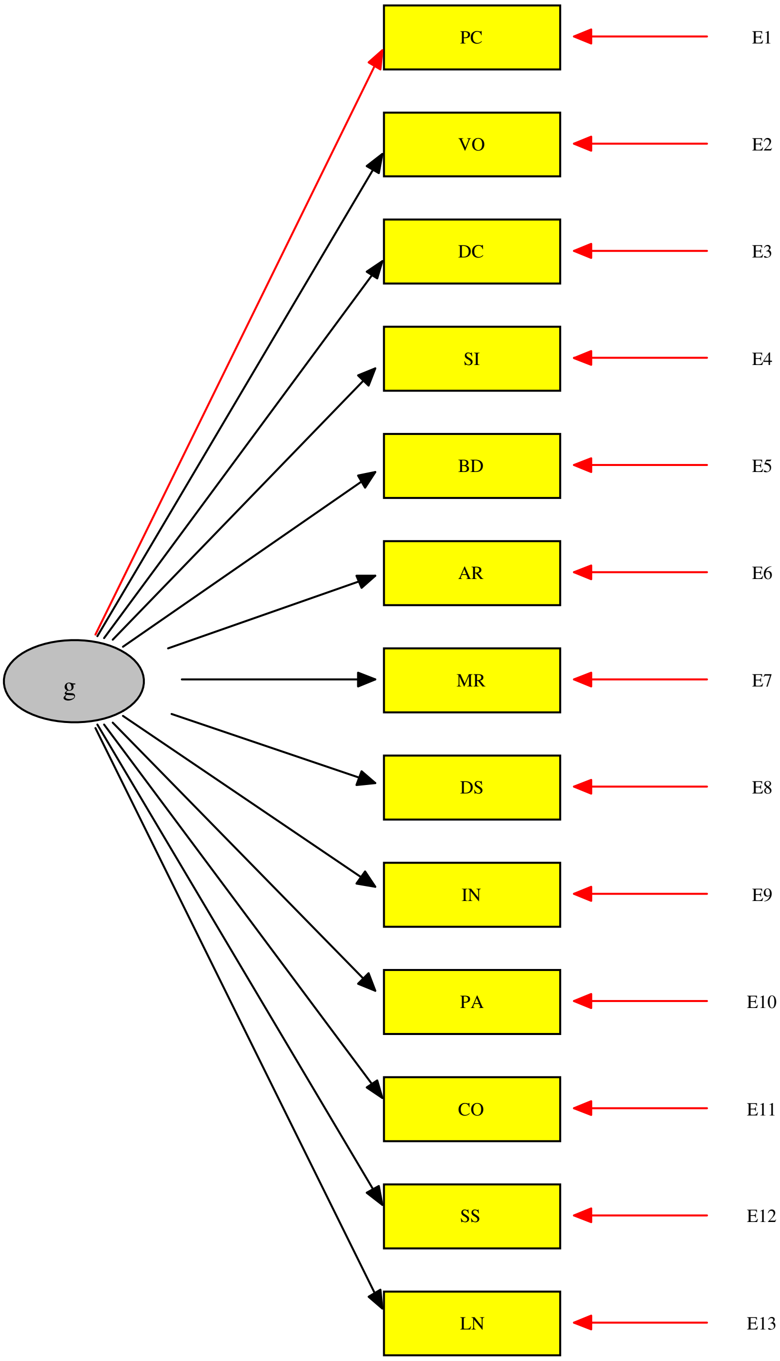
7.3 Models tested

Confirmatory factor analysis (CFA) via EQS 6.1 program (Bentler, 2007) was used to assess the fit of each sample of data to each model. The following models were tested.

7.3.1 One factor model

The one factor model asserts that all subtests are dependent on a single general factor. The subtests are: Vocabulary (VO), Similarities (SI), Arithmetic (AR), Digit Span (DS), Information (IN), Comprehension (CO), Letter-Number Sequencing (LN), Picture Completion (PC), Digit Symbol Coding (DC), Block Design (BD), Matrix Reasoning (MR), Symbol Search (SS), Picture Arrangement (PA). The unobserved variable *g* is represented in Figure 3 by an ellipse. The subtests are known as the observed variables and are represented by rectangles in the diagram. The observed variables are the subtests represented by rectangles. Unique factors are also represented on the path diagram (e.g. E1). The unique factors are assumed to be uncorrelated with each other and the common factors and can be considered the error variance of the factor. The arrows leading to the observed variables from the common factors are known as factor loadings.

Figure 3. One factor model



7.3.2 Implicit model (two factor model)

The implicit model asserts that performance on seven subtests depend on an unobserved variable called Verbal represented in Figure 4 by an ellipse. Six other subtests are assumed to depend on another unobserved variable called Performance, represented by the other ellipse in the path diagram. The verbal and performance factors are known as common factors and are assumed to be explained by a second order *g* factor, also represented by an ellipse.

7.3.3 Three factor model

A three factor model consistent with that tested by the WAIS III test publishers and the explicit WAIS-R factor structure was also tested for fit. The verbal comprehension factor is assumed to influence the following subtests: vocabulary, similarities, arithmetic, information and comprehension. The perceptual organisation factor is assumed to influence the subtests: picture completion, block design, matrix reasoning and picture arrangement. The third factor, attention is assumed to influence digit span, letter number sequencing, digit symbol coding and symbol search. The three indices are assumed to be explained by a second order *g* factor (Figure 5).

Figure 4. Implicit two factor model

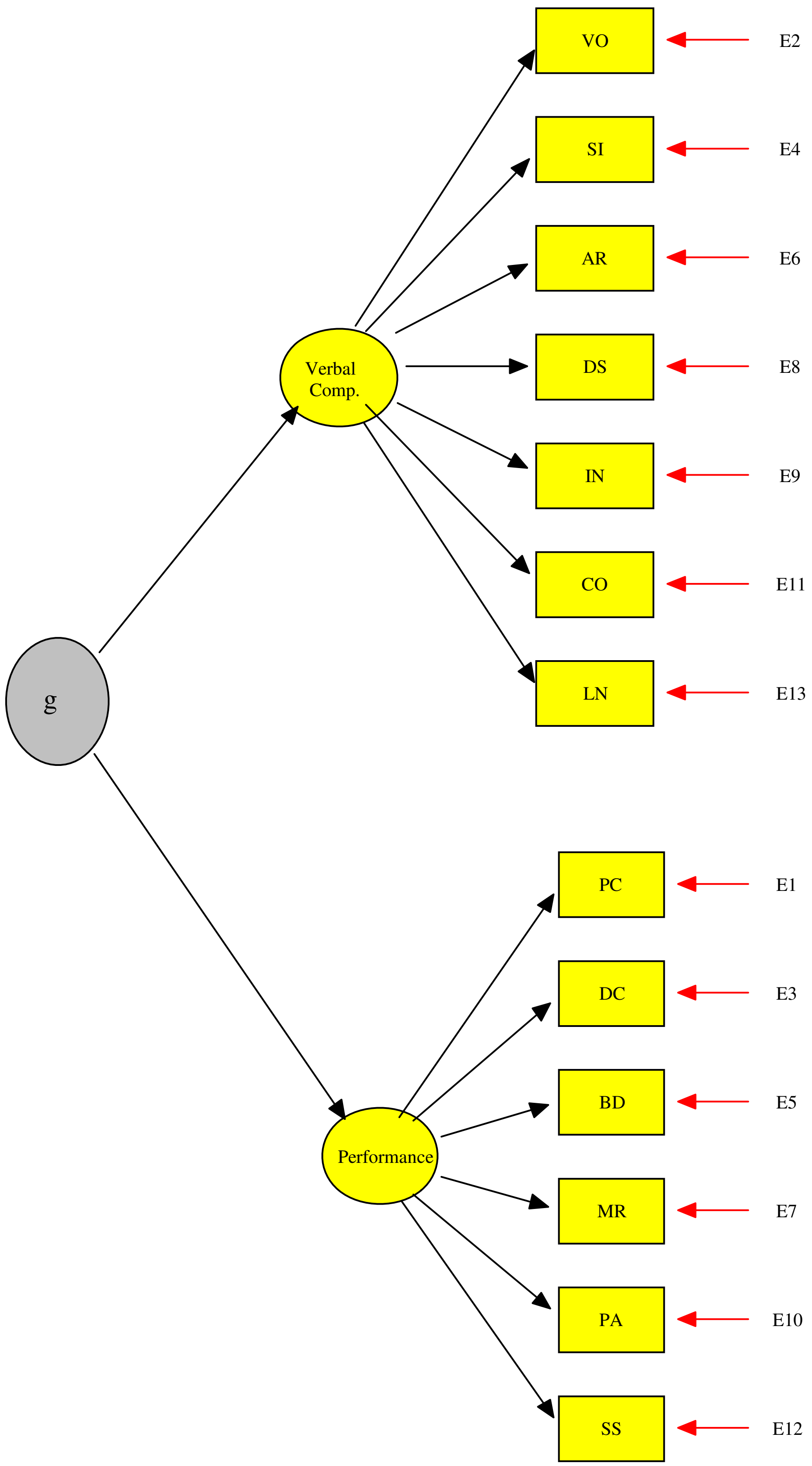
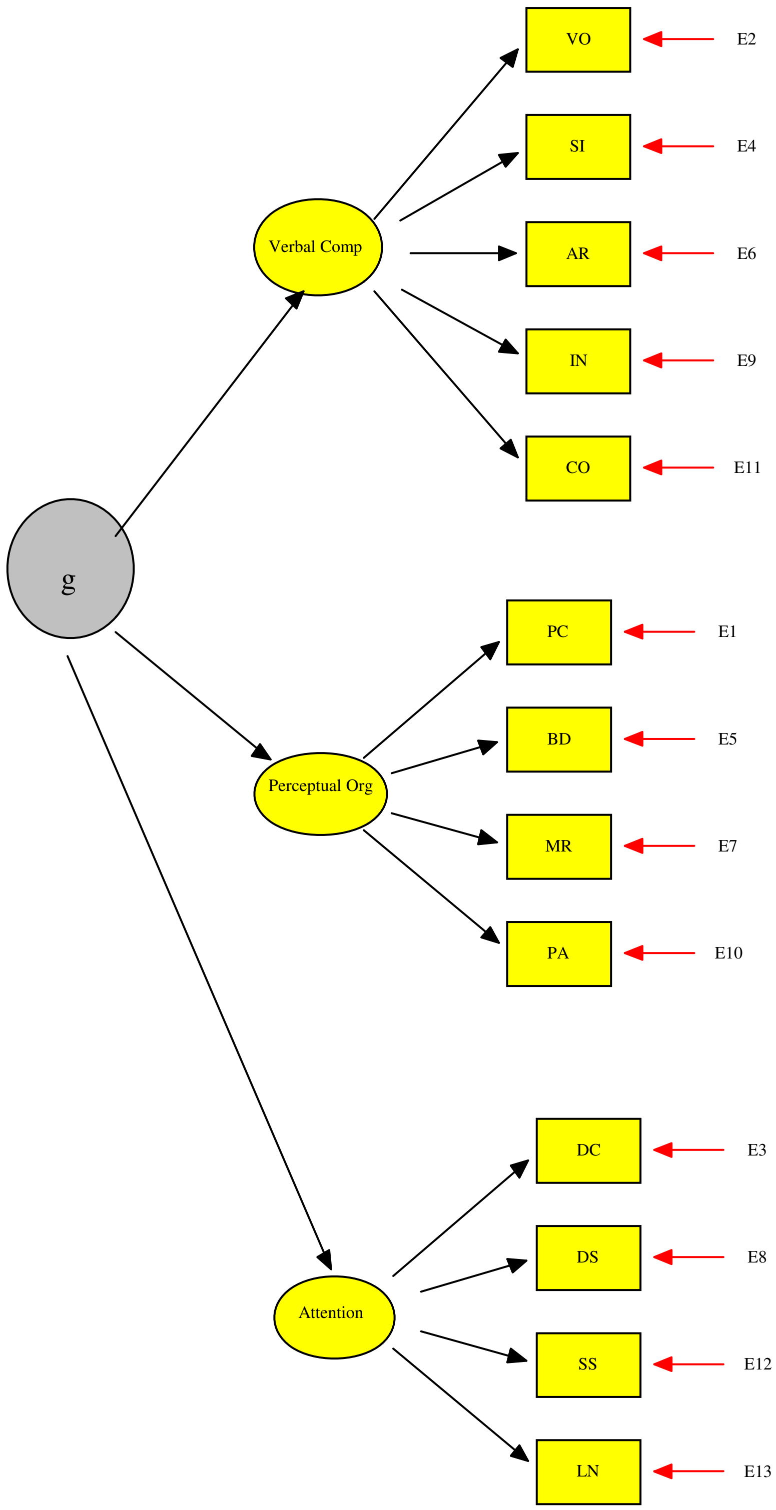


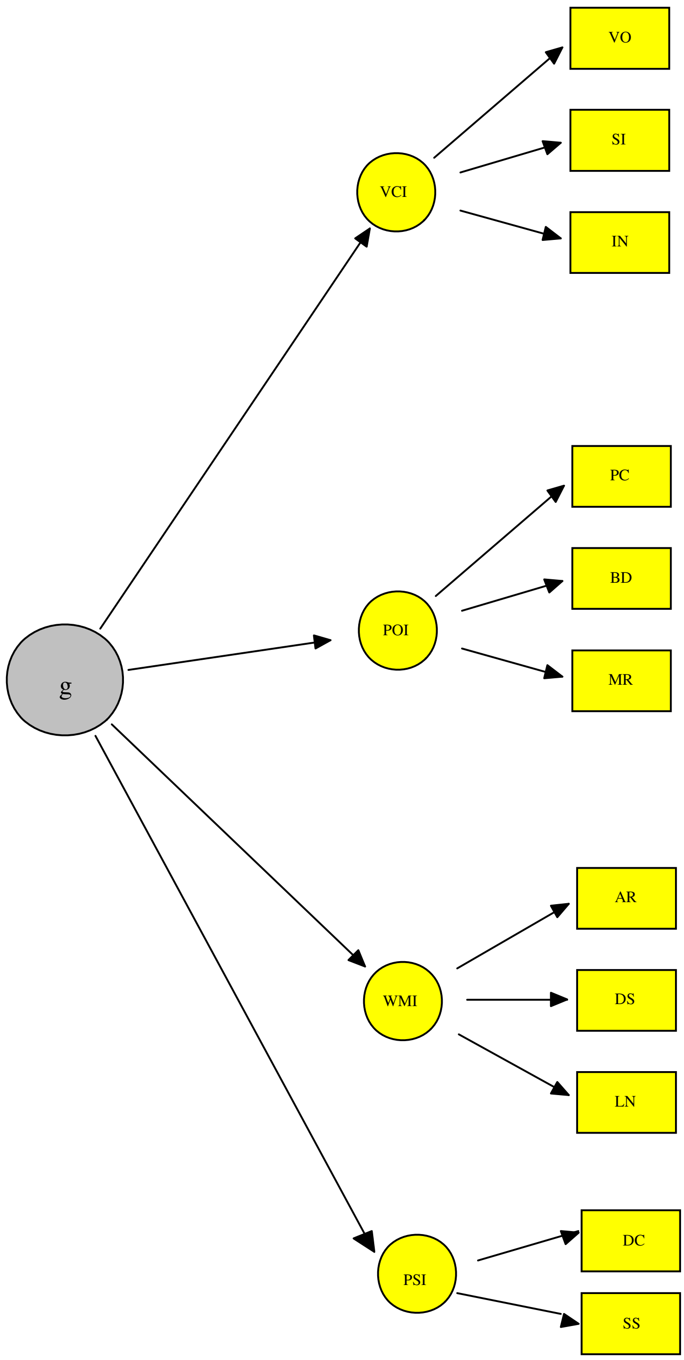
Figure 5 Three factor model



7.3.4 Explicit four factor model

In this explicit model there are four common factors assumed to explain performance on the subtests (see Figure 1). The Verbal Comprehension Index (VCI) is assumed to influence four subtests: vocabulary, similarities, comprehension and information. The Perceptual Organisation Index (POI) is assumed to influence four subtests: picture completion, block design, picture arrangement and matrix reasoning. The Processing Speed Index (PSI) is assumed to influence two subtests: digit symbol coding and symbol search. The Working Memory Index (WMI) is assumed to influence the three subtests: letter number sequencing, arithmetic and digit span. The four indices are assumed to be explained by a second order *g* factor. The explicit model was also tested using 11 subtests (excluding the two optional subtests of Comprehension and Picture Arrangement) which are sufficient for the calculation of index scores (Figure 6).

Figure 6. Four factor model (11 subtests)



7.3.5 Five factor CHC model

The five factor model is consistent with CHC theory and was found to show a better fit to the WISC-IV standardisation data (Keith *et al.* 2006) and the new WAIS-IV standardisation data (Keith *et al.* 2009). The five factors are Crystallised intelligence (Gc), Visual Reasoning (Gv), Fluid Intelligence (Gf), Short term memory (Gsm) and Processing Speed (Gs). Arithmetic is allowed to cross load on Gsm and Gf. The assumed relationship between these factors and subtests is shown in Figure 7.

7.3. 6 Six factor model

A six factor model consistent with updated *gf-gc* theory was also tested for fit based on findings by Burton *et al.* (2002). Six factors representing Semantic Memory, Verbal Reasoning, Perceptual Organisation, Working Memory and Processing Speed and the subtests assumed to correspond these factors are represented in Figure 8. In a study exploring the fit of more complex models to those tested by the WAIS –III publishers, this model was found to offer the best fit to a clinical sample with cross validation in the standardisation sample (Burton *et al.,* 2002). The model is similar to the CHC model but divides crystallised intelligence into two components: semantic memory and verbal reasoning. Arithmetic is allowed to cross load on Verbal Reasoning and Working Memory.

Figure 7. Five factor model

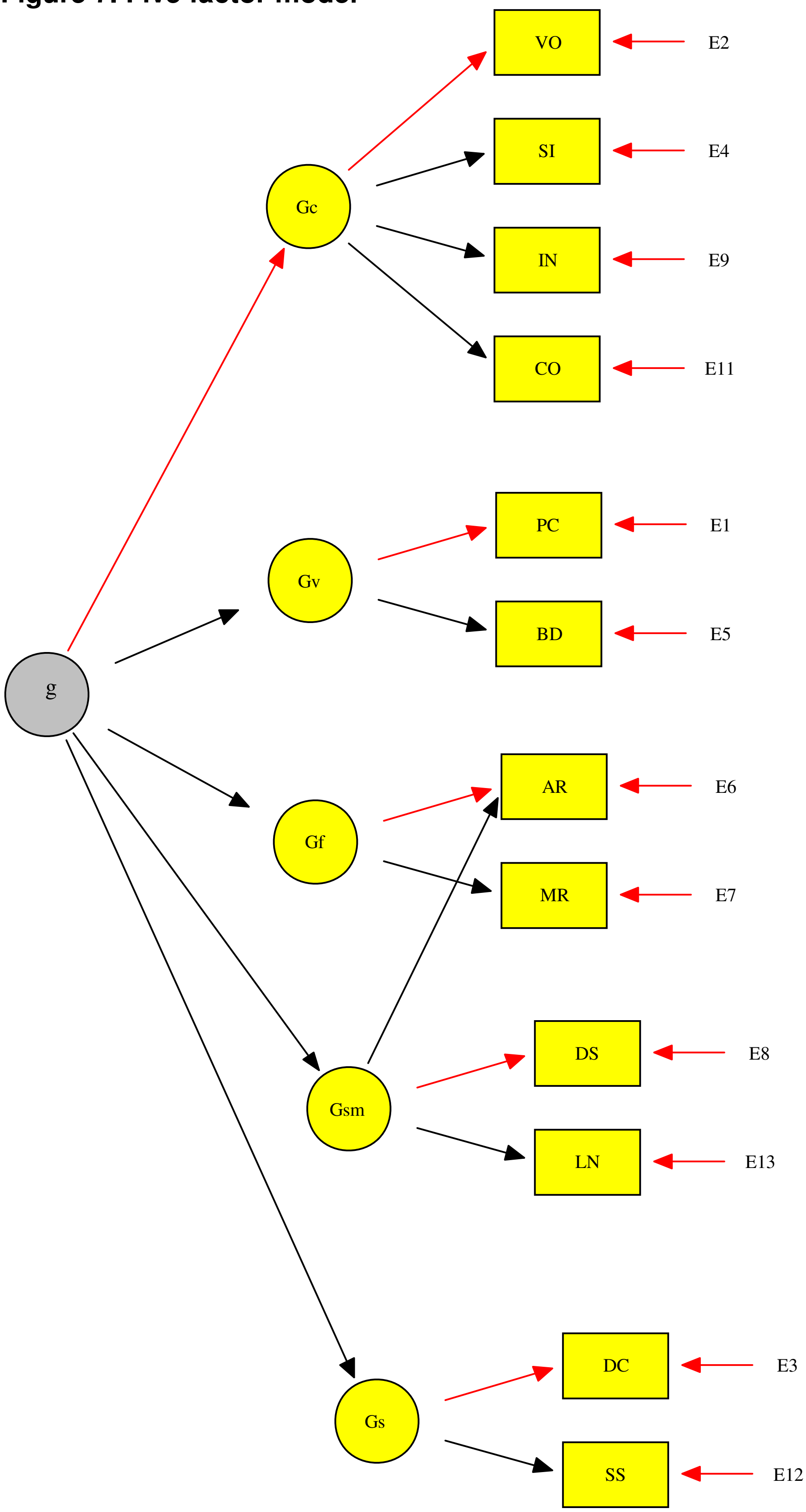
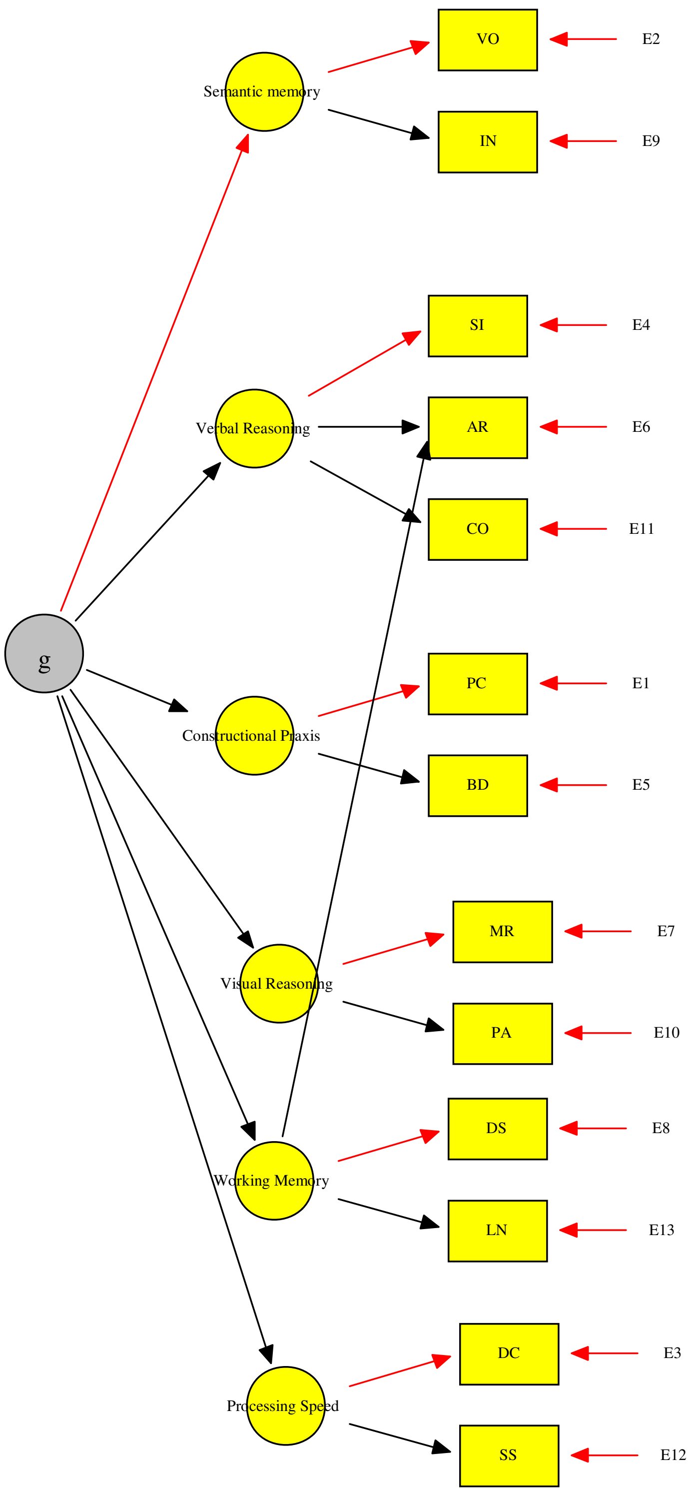


Figure 8. Six factor model



8. Results

8.1 Preliminary Analysis

A computer evaluation of assumptions using SPSS version 17.0 was carried out to assess the suitability of the data sets for analysis (Tabachnick & Fidell, 2001).

Descriptive statistics by subtest for each sample are provided in Tables 8, 9 and 10.

Table 8. Descriptive statistics by subtest for the sample with severe learning disability

Subtest	N	Minimum	Maximum	Mean	Std. Deviation
Picture Completion	140	1	6	2.26	1.173
Vocabulary	140	1	5	1.31	0.915
Digit Symbol Coding	140	1	5	1.68	0.807
Similarities	140	1	6	2.22	1.419
Block Design	140	1	6	2.46	1.266
Arithmetic	140	1	4	1.31	0.589
Matrix Reasoning	140	1	8	3.34	1.333
Digit Span	140	1	6	2.55	1.294
Information	140	1	6	2.70	1.044
Picture Arrangement	140	1	6	2.34	1.267
Comprehension	140	1	4	2.31	0.813
Symbol Search	140	1	5	1.34	0.794
Letter – number sequencing	140	1	5	1.32	0.742

Table 9. Descriptive statistics by subtest for the sample with significant learning disability

Subtest	N	Minimum	Maximum	Mean	Std. Deviation
Picture Completion	264	1	12	8.99	1.82
Vocabulary	264	1	9	7.47	1.36
Digit Symbol Coding	264	1	9	7.19	1.43
Similarities	264	1	9	9.09	1.84
Block Design	264	1	10	9.47	1.41
Arithmetic	264	1	11	6.02	1.62
Matrix Reasoning	264	1	9	8.89	1.16
Digit Span	264	1	11	9.36	1.54
Information	264	1	11	9.35	1.59
Picture Arrangement	264	1	9	9.01	1.57
Comprehension	264	1	8	7.23	1.24
Symbol Search	264	1	9	6.27	1.76
Letter – number sequencing	264	1	12	6.43	1.95

Table 10. Descriptive statistics by subtest for sample with borderline IQ

Subtest	N	Minimum	Maximum	Mean	Std. Deviation
Picture Completion	89	1	14	6.49	2.523
Vocabulary	89	2	10	5.40	1.593
Digit Symbol Coding	89	1	12	4.92	2.063
Similarities	89	1	10	6.12	1.338
Block Design	89	4	12	7.07	1.782
Arithmetic	89	1	10	5.46	1.949
Matrix Reasoning	89	2	12	6.72	1.989
Digit Span	89	2	10	6.12	1.744
Information	89	3	11	5.93	1.894
Picture Arrangement	89	2	14	6.66	2.210
Comprehension	89	2	14	5.12	2.245
Symbol Search	89	1	12	5.48	2.297
Letter Number Sequencing	89	1	10	5.66	2.369

8.2 Normality of Variables

Normality can be assessed through examination of histograms and by calculating standardised scores for skewness and kurtosis. Table 11 and 12 show the skewness and kurtosis z scores for each subtest in the three samples. Any value above 1.96 suggests the distribution of scores is non normal (Field, 2000).

Table 11. Skewness and Kurtosis z scores for each subtest for the learning disability samples

Subtest	Sample with severe learning disability Skewness z score	Sample with severe learning disability Skewness z score	Sample with significant learning disability Kurtosis z score	Sample with significant learning disability Kurtosis z score
Picture Completion	0.745	-0.281	0.959	**1.746**
Vocabulary	0.683	0.513	0.942	1.363
Digit Symbol Coding	1.318	**2.484**	1.080	**1.716**
Similarities	0.993	-0.019	-0.410	-0.49
Block Design	0.594	-0.186	0.461	0.49
Arithmetic	**1.940**	**3.683**	1.369	**3.269**
Matrix reasoning	0.248	0.661	0.456	1.195
Digit Span	0.808	0.206	1.010	1.817
Information	0.321	0.107	1.336	**2.270**
Comprehension	0.484	-0.790	0.159	-0.193
Picture Arrangement	0.277	-0.331	1.029	1.352
Symbol Search	**2.458**	**5.598**	0.651	-0.162
Letter Number Sequencing	**2.618**	**6.980**	1.164	1.812

Table 12. Skewness and Kurtosis z scores for each subtest for the borderline sample

Subtest	Sample with borderline IQ Skewness z score	Sample with borderline IQ Kurtosis z score
Picture Completion	.777	.652
Vocabulary	.767	.583
Digit Symbol Coding	1.365	**3.073**
Similarities	.710	**4.973**
Block Design	.151	-.343
Arithmetic	.246	-.510
Matrix reasoning	.388	-.340
Digit Span	.131	-.331
Information	.974	.153
Comprehension	.599	.425
Picture Arrangement	.753	.378
Symbol Search	.433	.834
Letter Number Sequencing	.148	-.630

Values in bold indicate those distributions which may depart from normality and supported the information from histograms which suggests that the distribution of some subtests is not normal. A Kolmogorov-Smirnov test suggested that all the subtests differed significantly from a normal distribution in all populations. Smith (2003) suggests that this may be expected given the limited range of scores from individuals with low IQ and that data distribution from a clinical population may not be expected to be normal (Tabachnick & Fidell, 2001). As such, it was considered appropriate to employ EQS 6.1 (Bentler, 2007), a non parametric test, to run the confirmatory factor analysis. Yuan and Bentler (1998) suggest, in fields such as psychology when data does not always meet assumptions of multivariate normality, it is not always appropriate to use normal theory methods, which can distort the results. Using the EQS, ROBUST option a Satorra-Bentler (1994) scaled test statistic corrects the mean of the sampling distribution so that it is closer to the expected mean. Studies have shown that this statistic works well for non normal and normal data (Curran *et al.* 1996) and is considered the best for dealing with non normal data (Bentler, 2007). A Bonnett-Woodward-Randall test in EQS showed significant excess kurtosis at the .05 level in the sample with borderline IQ and one case was removed. In the samples with learning disability multivariate kurtosis was not significant as the .05 level and no outliers were removed. The remaining data set for the sample with severe learning disability was n=140, significant learning disability was n= 264 and n=88 for the sample with borderline IQ.

8.3 Multicollinearity and Singularity

The determinant of the correlation matrix was 0.080 for the sample with severe learning disability, 0.106 for the sample with significant learning disability and 0.44 for the sample with borderline IQ. This indicates that variables were neither highly correlated (multicollinear) variables nor perfectly correlated (singular).

8.4 Factorability

An examination of the correlation matrix for the sample with severe and significant learning disability revealed 19% and 17% respectively of correlations in the matrix were at least 0.3. This suggests the data is factorable (Tabachnick & Fidell, 2001). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.607 and 0.711 respectively. Kaiser and Rice (1974) suggest values above 0.6 are considered good sampling adequacy for factor analysis (Field, 2000). For the sample with borderline IQ only 15% of correlations were >.3. The KMO measure of sampling adequacy was 0.566 which is considered acceptable but mediocre (Field, 2000).

9. Model estimation for the samples with learning disability

In models with data which is considered multivariate normal maximum likelihood estimation is typically used to estimate the models and in general a lower χ^2/df ratio, indicates a better fit. When data does not meet the conditions for multivariate normality a more valid approach is to use the Satorra-Bentler scaled χ^2 (S-B c^2). The Comparison Fit Index (CFI) is a value between 0 to 1.0; the closer the value to 1.0 the better the fit. The CFI provides an estimate of fit which takes account of possible misspecification caused by small samples. (Bentler, 1990) The root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) should be approximately .05 or less when indicative of a good fit. Both the CFI and RMSEA were adjusted using the Satorra-Bentler scaled χ^2 statistic to assess the models. Tables 13 and 14 show the fit statistics for the samples with severe and significant learning disability.

9.1 Results for the sample with severe learning disability

Table 13. Model estimation for the sample with severe learning disability

Fit Statistic	**S-B χ^2**	**df**	**p level**	**CFI**	**RMSEA** (90% confidence interval range)
One Factor	168.7	65	.00000	.557	.107 (.087-.127)
Two Factor	157.5	64	.00000	.601	.103 (-.082-.122)
Three factor	152.0	62	.00017	.616	.102 (.066 -.123)
Four factor (11 subtests)	87.9	40	0.0005	.746	.090 (.063-.116)
Four factor (13 subtests)	134.5	61	.00000	.687	.093 (.071-.114)
Five Factor CHC model	99.7	48	.00002	.752	.088 (.063 - .112)
Six factor	141.03	58	.00000	.645	.101 (.080-.122)

As can be seen from Table13 the best fitting models were the four factor (11 subtests) and the five factor model. However, none of the measurements of fit were indicative of good fit (non significant χ^2, CFI>0.9, RMSEA < .05).

Table 14. Model estimation for the sample with significant learning disability

Fit Statistic	**S-B χ^2**	**df**	**p level**	**CFI**	**RMSEA** (90% confidence interval range)
One Factor	312.3	65	.00000	.483	.120 (.107-.133)
Two Factor	172.1	64	.00000	.774	.080 (.066-.094)
Three factor	176.0	62	.0000	.762	.084 (.069 -.098)
Four factor (11 subtests)	86.6	40	0.0003	.869	.067 (.047-.085)
Four factor (13 subtests)	131.5	61	.00001	.853	.066 (.051-.082)
Five Factor CHC model	107.4	48	.00002	.856	.069 (.051 - .086)
Six factor	188.3	58	.00000	.728	.092 (.078-.107)

Similar to the data from the sample with severe learning disability, Table 14 indicates that none of the models showed a good fit (non significant χ^2, CFI>0.9, RMSEA<.05). The best fitting models are the four factor (11 and 13 subtests) and five factor model and at best these show a weak fit to the data with CFI nearing 0.9 and the lower end of the confidence limit for RMSEA around .05.

9.2 Interpreting factor loadings

A loading of .3 is equivalent to 9% of the variance in the indicator variable being explained by the factor and this is sometimes taken as the acceptable minimum level for factor analysis (Child, 2006). However, views on the requisite size of loading vary depending on the data, and some consider a loading of .6 high and those below .4 low (Hair *et al*., 1998). The squared multiple correlation is R^2 and is given alongside the factor loading for first and second order factors. R^2 expressed as a percentage shows how much of the variance is explained by the factor. Tables 15-22 show the two best fitting models for both samples. The remaining models can be found in Appendix 5.

Table 15. Standardised solutions and squared multiple correlations for a four factor model (11 subtests; sample of people with severe learning disability)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Multiple Squared Correlation
PC		.236			.056
VO	1.000				1.00
DC				.852	.725
SI	.367				.135
BD		.867			.753
AR			.608		.370
MR		.386			.149
DS			.687		.472
IN	.034				.001
SS				.467	.218
LN			.511		.261

Table 16. Loadings on second order factor and squared multiple correlations for a four factor model (11 subtests; sample of people with severe learning disability)

Factor	Loading on 2^{nd} order factor	Squared Multiple Correlation
VCI	.308	.095
POI	.702	.492
WMI	.709	.503
PSI	.745	.555

Table 17. Standardised solutions and squared multiple correlations for a five factor model (sample of people with severe learning disability)

Subtest	Loadings on Gc Factor	Loadings on Gv Factor	Loadings on Gf	Loadings on Gsm	Loadings on Gs	Squared Multiple Correlation
PC		.243				.059
VO	.651					.423
DC					.865	.748
SI	.461					.213
BD		1.000				1.00
AR			.516	1.000		.556
MR			-.449			.201
DS				.621		.386
IN	.281					.079
CO	.513					.263
SS					.460	.212
LN				.510		.260

Table 15 and 17 show that the minimum level of 9% variance explained was reached by all subtests except picture completion and information. Tables 16 and 18 suggest the VCI and Gc factors are least well explained by the second order g factor.

Table 18. Loadings on second order factor and squared multiple correlations for a five factor model (sample of people with severe learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Gc	.379	.143
Gv	.674	.454
Gf	-1.00	1.00
Gsm	.783	.613
Gs	.702	.492

Table 19. Standardised solutions and squared multiple correlations for a four factor factor model (11 subtests; sample of people with significant learning disability)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Squared Multiple Correlation
PC		.403			.162
VO	.715				.511
DC				.579	.336
SI	.515				.266
BD		.560			.313
AR			.422		.178
MR		.408			.166
DS			.451		.203
IN	.665				.443
SS				.793	.628
LN			.668		.446

Table 20. Loadings on second order factor and squared multiple correlations for a four factor model (11 subtests; sample of people with significant learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
VCI	.141	.020
POI	.914	.836
WMI	.506	.256
PSI	.839	.704

Table 21. Standardised solutions and squared multiple correlations for a five factor model (sample of people with significant learning disability)

Subtest	Loadings on Gc Factor	Loadings on Gv Factor	Loadings on Gf	Loadings on Gsm	Loadings on Gs	Squared Multiple Correlation
PC		.420				.177
VO	.682					.465
DC					.578	.334
SI	.553					.305
BD		.612				.375
AR			.177	.293		.153
MR			.533			.284
DS				.433		.187
IN	.669					.447
CO	.518					.269
SS					.794	.631
LN				.735		.540

Table 22. Loadings on second order factor and squared multiple correlations for a five factor model (sample of people with significant learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Gc	.108	.012
Gv	.837	.701
Gf	.752	.565
Gsm	.467	.218
Gs	.833	.694

Tables 19 and 21 show that the 9% minimum acceptable level of variance was explained by all subtests except arithmetic. Table 20 and 22 suggest that VCI and Gc factors are least well explained by a second order g factor.

9. 3 Results for the sample of people with borderline IQ

Due to the small size of the sample of people with borderline IQ the models did not converge for the second order analysis. First order analyses were run for each of the models and the results for the two best fitting models are shown in Tables 24 & 25. The remaining models can be seen in Appendix 5. Table 23 shows the model estimation for the sample with borderline IQ.

Table 23. Model estimation for the sample with borderline IQ

Fit Statistic	**S-B χ^2**	**df**	**p level**	**CFI**	**RMSEA** (90% confidence interval)
One Factor	161.6	55	.00000	.499	.131 (.105-. 155)
Two Factor	146.4	64	.00000	.572	.122 (.095-.147)
Three Factor	111.3	62	.00012	.744	.096 (.066-.123)
Four factor (11 subtests)	67.0	38	0.0026	.793	.094 (.055-.129)
Four factor (13 subtests)	85.7	59	.01318	.861	.072 (.034-.103)
Five Factor CHC Model	81.3	43	.00038	.803	.104 (.068-.137)
Six factor	78.6	49	.00466	.872	.086 (.048-.119)

Table 23 shows that the best fitting models are the four factor (13 subtests) and the six factor model. p values are higher indicating a better fit, CFI values are approaching 0.9 and the lower confidence limit for RMSEA is less than .05. Similar to the models for the severe and significant learning disability samples, none of the models meet the threshold for good fit.

Table 24. Standardised solutions and squared multiple correlations for a four factor model (13 subtests; sample of people with borderline IQ)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Squared Multiple Correlation
PC		.611			.374
VO	.968				.937
DC				.544	.296
SI	.435				.189
BD		.387			.150
AR			.357		.127
MR		.312			.097
DS			.772		.596
IN	.389				.151
PA		.559			.312
CO	.611				.373
SS				1.000	1.000
LN			.558		.312

Table 25. Standardised solutions and squared multiple correlations for a six factor model (sample of people with borderline IQ)

Subtest	Loadings on Semantic Memory Factor	Loadings on Verbal Reasoning Factor	Loadings on Constructional Praxis	Loadings on Visual Reasoning Factor	Loadings on Working Memory Factor	Loadings of Processing Speed Factor	Squared Multiple Correlation
PC			.402				.162
VO	.857						.734
DC						.546	.298
SI		.593					.352
BD			.288				.083
AR		-.344			.878		.230
MR				.443			.196
DS					.825		.680
IN	.397						.157
PA				.682			.465
CO		.580					.336
SS						1.000	1.000
LN					.557		.310

Tables 24 and 25 indicate the 9% minimum for acceptable level of variance explained for all subtests except block design.

In summary, for the sample with severe learning disability, the results suggest a poor fit for all models with the best fit found for a second order four factor model (11 subtests), CFI>.746 and five factor model CFI>.752. For the sample with significant learning disability the models demonstrated a weak fit to the data. The models showing best fit were the same as those showing best fit to the data from the sample with severe learning disability CFI > .869 for the four factor model and CFI> .856 for the five factor model. For the sample with borderline IQ, weak fit was found for the six factor model consistent with updated *gf-gc* theory (CFI=.872) and for the four factor model (13 subtests; CFI = .861).

10. Discussion

This chapter will discuss the results of this research and its implications with regard to the literature in the introductory chapter. The chapter will also examine the clinical and theoretical implications of the findings, the limitations of the research, and suggestions for future research.

The aim of this study was to examine the internal construct validity of the WAIS-III for people with severe and significant learning disability and those with borderline IQ. Confirmatory factor analysis was used to examine the fit of the data in these three populations to the implicit and explicit factor structures proposed by the test publishers in addition to more and less complex models. The research will also add to the discussion about the cut-off point used to define a learning disability and whether this reflects an actual difference in performance on these scales as measured by the WAIS-III. The results of this study can also be viewed in the light of the new WAIS-IV, its' theoretical underpinnings and the interpretative approach advocated by the test publishers.

10. 1 Results of the confirmatory factor analysis

A confirmatory factor analysis is a means of testing how the data fits models proposed by theory. The factor structure is suggested *a priori* and seeks to confirm a structure suggested by exploratory factor analysis. The present study was informed by previous research which variously provided support for the explicit four factor structure of the WAIS-III in non clinical populations (Garcia *et al.,* 2003; Saklofske *et al.* 2000; Taub, 2001; Arnau & Thompson, 2000) and in clinical populations

(Heijden & Donders, 2003; Ryan & Paulo, 2001; Dickinson *et al.,* 2002). Other research has found a six factor model (Burton *et al*. 2002) corresponding to updated *gf-gc* theory provided a better fit in both clinical and standardisation samples while research by Keith *et al.*(2006, 2009) suggests a five factor model consistent with CHC theory provides a better explanation of the data for the newer Wechsler adult and children's scales. Only one study was found which examined model fit in a sample of people with low IQ (Jones *et al.,* 2006) and this did not provide support for the explicit four factor model proposed by the test publishers (Psychological Corporation, 1997).

The results of the present study suggest the best fitting models for people with learning disability were four and five factor models. The best fit was found for the explicit four factor model using 11 subtests; the next closest fit was found for a five factor model, consistent with CHC theory; followed by the explicit four factor model using all 13 subtests. However, none of the models reached the threshold of CFI=.9, RMSEA <.05 which would indicate a good fit. Fit was weak for the sample with significant learning disability and poor for the sample with severe learning disability.

In a sample with borderline IQ, best fit was found for the explicit four factor model (13 subtests) and a six factor model consistent with *gf-gc* theory. None of these models also reached the threshold of CFI =.9, RMSEA <.05. The results for each group will be discussed in turn, followed by a discussion of the implications for the use of the WAIS-III in these populations.

10.2 Results for the samples with learning disability

The results from the present study did not find good support for any of the tested models in either learning disability sample. However, in both the severe and significant learning disability sample best fit out of all the models was for the four factor (11 subtests) model. This is in keeping with the findings by the test publishers (Psychological Corporation, 1997) supporting a four factor model and the results from studies using the standardisation sample and both clinical and non clinical samples (Garcia *et al.,* 2003; Saklofske *et al.* 2000; Taub, 2001; Arnau & Thompson, 2000; Heijden & Donders, 2003; Ryan & Paulo, 2001; Dickinson *et al.,* 2002). Compared to the less complex models, the four factor model (13 subtests) also provides a better fit for the data as does the Keith *et al.* (2006) five factor model which is consistent with CHC theory.

CHC theory's hierarchical framework proposes three strata of cognitive abilities: stratum III refers to *g* or general intelligence; stratum II refers to ten broad cognitive abilities, Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-term Memory (*Gsm*), Visual Processing (*Gv*), Auditory Processing (*Ga*), Long-term Retrieval (*Glr*), Processing Speed (*Gs*), and Decision/Reaction Time or Speed (*Gt*), Reading and Writing (*Grw*), and Quantitative Knowledge (*Gq*; McGrew & Flanagan, 1998) and stratum I refers to approximately 70 narrow cognitive abilities. The findings that the five factor model shows a better fit than the less complex one, two and three factor models is in keeping with the findings from Keith *et al*. (2006) in their analysis of the standardisation sample for the WISC-IV and preliminary analysis of the new WAIS-IV by Keith (2009) cited in Lichtenberger & Kaufman, (2009). The CHC model tested contains a Crystallised Knowledge (Gc) factor measured by the same four subtests used to calculate the VCI index; a Visual Processing (Gv) factor measured by picture completion and block design; a Fluid Reasoning (Gf) factor measured by matrix reasoning and arithmetic; a Short term memory (Gsm) factor measured by digit span and letter number sequencing and arithmetic; and a Processing Speed (Gs) factor measured by digit symbol coding and symbol search. Thus the main difference between the four and five factor model is POI is divided into two components; visual processing and fluid reasoning with arithmetic free to load on both fluid reasoning and short-term memory. Loadings suggest the Gsm index betters explains the variance in the arithmetic subtest than the Gf factor in the sample with a learning disability.

In the version of six factor *gf-gc* theory tested here the subtests of similarities and comprehension are thought to load on a fluid reasoning factor rather than a crystallised intelligence factor as proposed by the five factor model. The better fitting four and five factor model over the six factor model may suggest the four subtests of vocabulary, comprehension, similarities and information are more appropriately grouped together for some individuals with a learning disability.

Despite this trend towards support for a four and five factor model, it should be emphasised that fit could only be described as weak in the sample with significant learning disability and poor in the sample with severe learning disability. The CFI did not meet the threshold for good model fit (0.9) and RMSEA was not lower than

.05 except at the lower end of the confidence interval for one model (four factor 11 subtests) in the sample with significant learning disability..

In some models a large proportion of the variance in the factors could be explained by the second order *g* factor. This was found, for example, in the sample with significant learning disability, in the six factor model on the Gv factor comprising picture completion and block design and the Gf factor comprising matrix reasoning and picture arrangement (Table 43). This is replicated in most of the other performance subtests across models in the sample with significant and severe learning disability. Verbal tasks seem less well accounted for by a second order *g* factor. Research by Taub (2001) found the performance factor was completely subsumed by the second order factor consistent with *g* leading the authors to question the validity of the PIQ factor distinct from a general intelligence factor. The findings of this research also suggest in individuals with significant learning disability the variance accounted for by PIQ/ POI/ Fluid reasoning factors are to a great degree explained by second order *g*.

CHC theory and updated *gf-gc* theory propose a distinction between crystallised intelligence (*gc*) tasks and those considered to be dependent on education and experience and fluid intelligence tasks (*gf*), thought to be linked to the efficiency of internal mechanisms (Cattell,1963). The verbal/VCI/ *gc* components of the tasks are largely influenced by educational experiences which may be varied in a population of individuals with a learning disability (DoH, 2001). Recent work by Ryan *et al.* (2002) suggests that performance subtests on the WAIS-III consistent with fluid intelligence tasks are more vulnerable to deterioration across the life span whereas the verbal/crystallised tasks remain stable. Ardila (2007) found increased heterogeneity in subtests requiring executive functions, attention and some non-verbal abilities whereas a more stable homogeneous pattern of decline was suggested for visuo-constructive abilities and general knowledge. If the crystallised components of intelligence are represented by the VIQ, VCI, *gc* elements of the Wechlser scales, they highlight the 'learned' aspects of our intelligence and those most likely to show a stable decline across the life span compared to less predictable decline in *gf* skills. Compared to fluid reasoning skills they also appear to be explained less well by a general intelligence factor. This differentiation in skills would also explain the poorer fit of a one factor model for the samples with learning disability.

10.3 Implications for interpretation of the WAIS-III for individuals with learning disability

The results of this study showed little support for the explicit or implicit construct validity of the WAIS-III in a population with severe learning disability (IQ< 55). Support was not good for less complex models for the population with significant learning disability however a four factor and five factor model showed the best fit to the data. In their confirmatory factor analysis of the WAIS-III in a clinical sample, cross validated with the standardization sample, Burton and colleagues (2002) found a six factor model, corresponding to an expanded version of *gf-gc* theory provided the best fit to the data and was more accurate than the test publisher's model in explaining the latent variability among subtests. However, Burton *et al.* (2002)

suggest a hierarchical view can accommodate this difference in findings. They suggest there is enhanced clinical utility for individuals in interpreting these results taking into account broad and narrow abilities and as such the interpretation of the WAIS-III will be most relevant to the individual being tested. An individual whose scores on all 13 subtests do not differ significantly from the mean score should allow accurate interpretation of FSIQ. Similarly, when scores for the subtests relating to verbal IQ and performance IQ do not differ greatly from their respective mean score, VIQ and PIQ can be seen as an accurate reflection of verbal abilities and performance abilities. A similar rationale would allow for interpretation across the four indices when there is limited variability across subtests. However, should there be greater variability across the subtests, indices may not accurately represent a unitary construct and there may be more clinical utility in interpreting scores based on a six factor model. While none of the models provided a good fit to the data the results of this study support the findings that theoretical models (from two to six factors) better explain the data from a clinical sample of people with a learning disability than a one factor model.

The findings from this study contrast with Jones *et al.* (2006) who found support for the implicit two factor solution but not for the explicit four factor solution in a low IQ population. A possible reason for this discrepancy in results is the difference in the sample used. The Jones *et al.* (2006) study excluded individuals with epilepsy, autistic spectrum disorders and those with brain injury in order to avoid the influence of scores from individuals whose scores may confound the results of a generic low IQ population. In the present study it was decided to use a clinical population so that the results would be more directly representative of the population usually seen by clinical psychologists. This rationale arguably increased the heterogeneity of the sample. Kline (1994) suggests that heterogeneous samples increase the variance and increase the loadings on factors and show more face validity when looking across a whole population. Thus, the findings from Jones *et al.,* (2006) may provide an accurate representation of a low IQ population which is more homogeneous in nature and thus loadings on factors are naturally reduced. This explanation of the Jones *et al.* (2006) data is consistent with the multi model interpretative approach outlined above. Where variability in scores is limited and individuals do not show large inter-subtest scatter, a two factor VIQ and PIQ distinction is a more useful model.

10.4 Results for the sample with borderline IQ

The results for the sample with borderline IQ show the best fitting models to be the four factor model (13 subtests) and a six factor model consistent with updated *gf-gc* theory. These findings are consistent with the explicit structure suggested by the test publishers (Psychological Corporation, 1997) and those of Burton *et al.* (2002) who found a six factor model comprising the following factors: Semantic Memory (Gc), Verbal Reasoning (Gf), Visual Reasoning (Gf), Working Memory (SAR), Constructional Praxis (Gv) and Processing Speed (Gs) provided the best fit to the data.

However, when examining the Satorra Bentler χ^2 (S-B- χ^2) statistic and other estimations of model fit it can be seen that a four factor model (13 subtests; CFI = .861, $p < .01318$, RMSEA 90% confidence interval =.034-1.03) and also the six factor model (CFI= .872, p<.00466, RMSEA 90% confidence interval =.048-.199) do not meet the threshold for a good fit to the data.

Like the learning disability sample, no exclusion criteria were applied to this population and so it is likely the sample is heterogeneous in nature and therefore possible that factors which might otherwise be masked by homogeneity become more apparent. Aside from the five factor model, fit did improve as the number of factors increased.

10.5 Implications for the interpretation of results for individuals with borderline IQ

The results of the current study suggest individuals with a borderline IQ may show a more varied subtest scatter than individuals with lower IQ levels. This highlights the possibility of using a more complex six factor model of interpretation consistent with updated *gf-gc* theory to interpret the results as well as the test publishers four factor model. The proposed structure of the updated version of *gf-gc* theory used in this study and by Burton *et al.* (2002) is similar to the revised CHC theory tested in this study. In updated *gf-gc* theory, *gc* is divided into two components known as semantic memory and verbal reasoning and this increased differentiation of the *gc* factor appears to fit better with the sample with borderline IQ. Semantic memory (comprising vocabulary and information) can be viewed as a facet of crystallised intelligence and the acquired learning of skills and knowledge. Verbal Reasoning (comprising similarities and comprehension) requires more fluid intelligence skills such as active problem solving. Therefore the finding that individuals with borderline

IQ show this difference in verbal skills lends greater support for interpreting the WAIS-III using a six factor model when inter subtest scatter is greater. In the sample of people with significant learning disability a four and five factor model was preferred to a six factor model and on this basis it could be suggested verbal skills are less differentiated in some individuals with a learning disability.

The results from the sample with borderline IQ do not provide such convincing support for the five factor approach consistent with CHC theory and findings by Keith *et al.* (2006, 2009) on newer versions of the Wechsler scales. The model's poor fit to the data may be a result of the low loadings of matrix reasoning and arithmetic on the Gf factor. In both five and six factor models, loadings for arithmetic were higher on the working memory factor and suggest in both populations arithmetic is better accounted for by memory rather than fluid or verbal reasoning factors. This would suggest in the CHC model, in the sample with borderline IQ, matrix reasoning alone is not a sufficient indicator of Gf and is a stronger indicator when combined with picture arrangement as in the six factor model.

The findings do not provide convincing support for a four factor (11 subtests) model or a one, two or three factor model. This questions the validity of the test when all 13 subtests are not completed. It also questions the implicit two factor model suggested by the test publishers (Psychological Corporation, 1997) and findings from a low IQ population (Jones *et al.,* 2006). However, this may be an artefact of the smaller sample size compared to the samples with learning disability. This will be discussed further in the Limitations section.

10.6 Further discussion of results

The results also suggest that symbol search is a strong indicator of processing speed and that vocabulary is a strong indicator of verbal comprehension or crystallised intelligence in individuals with borderline IQ. Lichtenberger & Kaufman (2009) suggest an individual's ability to process information rapidly is "dynamically related to one's ability to perform higher order cognitive tasks" (pp.20-21). Therefore we may expect a stronger correlation between processing speed and second order *g* compared to vocabulary which is more dependent on education and experiences. This was the finding in the second order analyses of the sample with learning disability. However, the results from the sample with a learning disability suggest an even stronger influence on *g* on the perceptual organisation factor and on subtests which appear to require skills of fluid intelligence. In the present study, due to the smaller sample size it was not possible to carry out a second order factor analysis of the data from the borderline IQ sample and examine how much of the variance present in the first order factors was explained by the second order factors. It would be interesting to explore whether *g* exerts a similar influence in a larger borderline sample.

The results of the current study show three of the four highest scores were achieved on POI subtests for the learning disability group. This was replicated in the borderline sample with all four POI subtests being the highest scoring. The lowest scoring subtests for the learning disability group were arithmetic, digit symbol coding, letter number sequencing and symbol search (consistent with the WMI and PSI indexes). The lowest scoring subtests for the borderline group were digit symbol coding, vocabulary, comprehension and symbol search for the borderline group (consistent with VCI and PSI indexes). Thus across all samples, individuals tend to perform better on POI subtests and poorest on PSI subtests; suggesting that individuals on either side of the cut-off struggle more with tasks requiring speed, and in general perform better on non verbal tasks compared to verbal tasks.

11. Ethical Implications for the field of learning disability and future research

11.1 Cut-off points

In order to meet the criteria for a diagnosis of a learning disability, an individual would typically show a full scale IQ score of less than 70. This corresponds to two standard deviations from the mean score for the normal population and represents a score above which 95% of the population would be expected to fall. Applying this strict cut-off point allows service providers to tailor their provision to those in the bottom 5% who could be expected to require the greatest levels of support. However, relying on this cut-off point places great importance on the validity and reliability of the tool used to assess IQ and arguably assumes those with borderline IQ scores may have a profile different to those falling below the cut-off. In general the findings from this study suggest a four factor model provided the best fit to both samples and as such the cut-off point is not indicative of an actual difference in cognitive profile. For the samples of people with learning disability fit was better for four and five factor models compared to four and six factor models for the borderline sample.

An arbitrary cut-off point arguably strengthens the need for psychologists to use informed clinical judgement to assess the need of an individual who may require the additional support provided by learning disability services. This will be particularly relevant to individuals who have a FSIQ just above or below the cut-off point, given the standard errors of measurement inherent in the Wechsler scales. Assuming the other criteria of impairment in adaptive functioning and onset before adulthood were met, clinicians would then have to make a decision as to whether a diagnosis of learning disability would be made. Strict interpretation of the criteria could result in individuals missing out on services, accessing support or lead to restrictions in parenting (Aunos *et al.,* 2005), In more extreme cases, in the USA for example, the diagnosis of a learning disability may save someone from the death penalty (Kanaya *et al.,* 2003).

This research did not find a difference in cognitive profile between those with IQ above and below 70 as measured by the WAIS-III and as a result would suggest the rationale for a cut-off is limited to the degree of impairment rather than a qualitative difference in profile.

In the sample of people with IQ under 70, sixty-five per cent met the criteria for a significant intellectual impairment and thirty-five per cent for severe intellectual impairment. While the trend of best fitting models is the same above and below IQ 55, support is not good for either the population with significant learning disability or those with severe learning disability. In this study, the mean of each subtest was at least a scaled score of two, however, scaled scores of one were a feature of the range of scores for all subtests, particularly those in the severe learning disability sample. Whitaker, (2005) suggests clinicians should be wary of interpreting IQ or index scores based on scaled scores of one as they do not differentiate scores more than three standard deviations below the mean. The results of this study echo Whitaker's findings. As yet there is no alternative measure of cognitive ability which can be used in the lower IQ ranges and the design of suitable tests would pose significant challenges. A relevant test of cognitive abilities for people with learning disability would have to take into account the ability of this population to understand verbal instruction or the testing procedure would be unable to differentiate cognitive constructs from verbal capability (McKenzie *et al.,* 2004).[4]

11.2 Using the WAIS-III with people with a learning disability

The results of this study provide weak support for the construct validity of the WAIS-III in a population with significant learning disability. There are also strong caveats which must be acknowledged when using the WAIS-III with this population. Findings suggest that practitioners do not follow the standardised instructions when administering the test to people with a learning disability and the long administration times may be considered inappropriate for individuals with attention difficulties or

[4] A similar debate on cut-off scores is to be found at the other end of the intelligence spectrum. Gifted cut-off rates are also recognised as problematic and it has been argued they should be seen within the context of test appropriateness given an individual's cultural background, the impact of speed on a set of scores, ceiling effects and the unique subtest profile (Kaufman & Lichtenberger, 2006).

those who are anxious about the testing procedure (McKenzie *et al.* 2004). Modifications to the standardised instructions may also inflate IQ scores (Joncas & Standing, 1988; Slate *et al.* 1991). Standardised and validated short forms such as the WASI address some of these concerns, however, some research suggests the test would misdiagnose some individuals (McKenzie & Paxton, 2006). Research on common modifications to the WAIS-IV will be an interesting addition to the literature.

In addition, it has been argued the low IQ sample used in the standardisation process for the WAIS-III is not representative of actual clinical populations (Murray *et al.* 2003). Whitaker (2008) highlights two major difficulties with the use of the WAIS-III with individuals with a learning disability. He argues first, the assumption of a normal distribution in the population potentially underestimates the numbers of people with a severe or profound learning disability and second, the floor of the test fails to discriminate those falling more than three standard deviations from the mean score. The results of this study suggest the implicit or explicit models suggested by the test publishers do not adequately explain scores in people with severe learning disability (IQ between 45- floor of test and 55).

11.3 The Flynn effect

The Flynn Effect can also have a potential impact on diagnosis of a learning disability. Most noticeable differences have been found at the beginning and end of norming cycles such that an individual who gained an IQ score over 70 in an old test may find their score on a new test is below 70 and is therefore eligible for services previously denied. It is possible that there may have been an effect of IQ drift in this study as data was collected from the start and end of the norming cycle for the WAIS-III. This could have resulted in higher IQ scores for participants completing the test in the later years. Flynn (2007) suggests that IQ is raised by about 0.3 of a point per year though it is unclear if this increase is still apparent particularly for people in the low IQ range (Whitaker, 2010a). An adjustment on this basis of a 0.3 point per year increase for full scale IQ scores was not carried out in this study but could be the focus of future studies. Had such an adjustment been made some individuals with borderline IQ scores may more accurately have been placed in the significant learning disability category while some from the lower end of the significant learning disability category may have been more accurately placed in the severe learning disability category. Research has begun to assess the contribution of the Flynn Effect to differences between scores on the old WAIS-III scale and the new WAIS-IV scale and initial research suggests the new normative data does make the test harder (Lichtenberger & Kaufman, 2009). Lower scores on the WAIS-IV could result in an increase in diagnoses of learning disability and it will be interesting to see whether, even with a diagnosis, individual support is safeguarded during times of national economic distress.

Whitaker (2010b) also suggests that sources of error in addition to the Flynn Effect should be considered. It is suggested in order to find a true 95% confidence interval for IQ score the error arising from a lack of internal consistency must be combined with the error arising from the lack of stability in the test. When these are combined with the Flynn effect Whitaker (2010b) suggests only an IQ below 42 would ensure the person had an IQ less than 70. This reflects a 95% confidence interval which extends to 28 points below the measured IQ score and below the floor for the test. To be sure of an IQ above 70 the measured IQ would need to be at least 88. These findings further question our ability to measure low IQ and the utility of current tests.

Recent research by Gordon *et al.* (2010) has raised further questions by indicating the WAIS-III produces higher scores than the WISC-IV. Both tests are used with individuals with a learning disability and using one test rather than the other may impact the individual and whether they receive services. The findings by Whitaker (2010a+b) suggest the reliability and validity of these tools in this population deserve more attention and in particular, child and adult services may need to collaborate more closely at times of transition when different testing protocols may lead to differing views on provision of services.

11. 4 Intelligence testing in individuals with learning disability

Theories of intelligence are sometimes delineated according to their internal validity or external validity. Contextual models emphasise the importance of an individual's adaptation to their environment and focus less on performance on testing procedures in artificial environments (Berry & Irvine, 1986; Gardner, 1983; Sternberg, 1985). The Wechsler scales provide a measurement of intelligence focused on a set of mental tasks which are assumed to underpin intelligence across the whole population. This study provides little support for the psychometric credentials of the WAIS-III in a population with intellectual impairment and the work by Whitaker

(2010b) casts doubts on the validity of scores across both significant and severe levels of impairment. The inclusion of measures of adaptive and social functioning in the diagnostic criteria for a learning disability has placed less emphasis on IQ scores (BPS, 2001). However, it is a financial reality that there will need to be some way for society to apportion limited services. Measures such as the Vineland Adaptive Functioning Scale (Sparrow *et al.* 1984) and the Adaptive Behaviour Scale (Harrison & Oakland, 2000) provide a good adjunct to the IQ test, however, their more subjective nature may be less well regarded than the psychometrically rigorous and objective IQ test, despite the well documented short-comings of the Wechlser scales when used with people with low IQ (Whitaker, 2005; McKenzie *et al.* 2004). Several authors have acknowledged the need for more appropriate means of testing intelligence in a low IQ population but currently there is no credible alternative to the Wechsler scales (Kaufman, 2000; Whitaker, 2005). Further research may help clinicians assess whether an alternative test can provide more ecologically relevant information about the degree of intellectual impairment and the type of support required.

11.5 Implications of the results for the WAIS-IV

The latest Wechsler adult intelligence scale, the WAIS-IV (Psychological Corporation, 2008) is proposed as a better fit with current theory, in particular developments in cognitive neuroscience and cognitive development in the areas of fluid reasoning, working memory and processing speed (Lichtenberger & Kaufman, 2009). It is no longer possible to extract a verbal and performance IQ and instead full scale IQ is calculated from the sum of the four indexes. A Global Ability Index

(GAI) can be calculated which sums the scaled scores on three verbal comprehension subtests (VCI; similarities, vocabulary, information) and three perceptual reasoning subtests (PRI; block design, matrix reasoning and, a new subtest, visual puzzles).

The GAI gives an estimate of general intellectual ability when the FSIQ is not interpretable, i.e. when the size of the difference between the index scores is equal to or more that 1.5 standard deviations (> 23 points) from the mean but the difference between the VCI and PRI indices is equal to or smaller than 1.5 standard deviations (<23 points) from the mean. Thus, for an individual with variability greater than 23 points between index scores, a GAI should be calculated provided there is less variability between their VCI and PRI scores. When VCI and PRI scores equal or exceed 1.5 standard deviations from the mean, neither the FSIQ or the GAI is interpretable and therefore it can be concluded that the individual's intellectual performance is best understood by reference to the four indices (VCI, PRI, WMI, PSI).[5] .

To align the scale with current theory, it is possible to interpret the WAIS-IV according to CHC five factor model (Keith, 2009). This provides clinicians an option to differentiate skills of fluid reasoning and visual processing over and above the aggregate perceptual reasoning score. The five factors and their corresponding

[5] Lichtenberger & Kaufman (2009) suggest the exception to this rule is when a global score is necessary for diagnosis of intellectual ability, be that a learning disability or placement in a programme for those considered gifted. In this case, clinical judgement is to be used to assess which global score most accurately represents the person's intellectual ability. In the example of an individual who is impulsive and distractible, the GAI may provide a more accurate reflection of ability as it does not include WMI and PSI (Lichtenberger & Kaufman, 2009). However, it is important to note the GAI may be higher than the FSIQ in individuals with a mild learning disability and 39% of individuals with a moderate learning disability showed a discrepancy of 5 or more points in favour of the GAI (Psychological Corporation, 2008).

subtests are Crystallized intelligence (Gc; vocabulary + information), Short-term memory (Gsm; digit span + letter-number sequencing), Fluid reasoning (Gf; matrix reasoning + figure weights), Visual Processing (Gv; block Design + visual puzzles) and Processing Speed (Gs; symbol search + coding). While the terminology is consistent with CHC theory, it is also possible to interpret the five factor model from a Lurian neuropsychological perspective. This perspective promotes the idea of three functional systems of the brain which relate to arousal and attention; sensory integration and storage and executive functioning. Lichtenberger and Kaufman (2009) suggest clinicians can use their own theoretical persuasion to inform their interpretation of the WAIS-IV scores be this from a CHC, Lurian or Wechsler tradition.

The findings from the present study show some support for the new direction taken by the WAIS-IV. A multi-model approach (or cross battery approach as described by Lichtenberger & Kaufman, 2009) which takes account of the individual's profile and allows the clinician to draw on relevant theory which best fits the data, be this based on CHC theory, *gf-gc* theory, Wechsler or other approach, is arguably more clinically relevant. Theorists may find this approach less satisfactory as several competing and different theories may be used to describe the same set of data. However, as Wechsler himself intended, the scales are first and foremost a clinical tool. As Burton *et al.* (2002) suggest:

" the publishers and developers of the Wechsler scales have traditionally focused on assessing interindividual differences in performance rather than attempting to measure intraindividual differences. Thus the focus of validation studies have been to assess the measurement efficiency of global indices rather than to derive more complex models that would lend themselves to the analysis of individual differences." (p.384).

For clinical psychologists, their use depends on their clinical utility and relevance. Research on the construct validity of the new scales has already begun (Canivez & Watkins, 2010) and it remains to be seen if clinical psychologists embrace the new tool in keeping with Wechsler traditions or a cross battery approach which permits interpretation of the data from an alternative theoretical perspective.

12. **Limitations of the study**

12.1 Sample size

There are no clear guidelines on the size of a sample suitable for factor analysis. Recommendations vary from $N-n-1 \geq 50$ (where N = number of participants and n = number of variables (Lawley & Maxwell, 1971), N at least 100 (Gorsuch, 1983), subject to variables ratio no lower than 5 (Bryant & Yarnold, 1995) to minimum desirable N =250 (Cattell, 1978). Comrey and Lee (1992) suggest a rating scale where 100 = poor, 200 = fair, 300 = good, 500= very good, 1000 or more = excellent. Costello and Osborne (2005) suggest when a ratio of subjects to variables is 2:1, only 10% of samples yielded correct solutions. When subject to variable ratios are 20:1 the yield rises to 70%. In the present study, the criteria of $N-n-1 \geq 50$ was met for both samples. The subject to variable ratio in the severe learning disability sample was 11:1, 20:1 in the significant learning disability sample and 7:1 in the borderline sample. It is worth noting sample size in Jones *et al.* (2006) of n=105 is equivalent to a subject to variable ratio of 8:1 and according to Costello and Osborne's (2005) data, this would only be expected to yield a correct solution in 50%

of samples and this questions the veracity of the results from the Jones *et al.* (2006) study and the borderline sample in the current study.

12.2 Sample

As can be expected from a clinical sample, descriptive analysis of the data sets demonstrated that the samples did not meet all conditions of normality or linearity. Normal theory methods may not provide accurate solutions when conditions of multivariate normality are not met (Yuan & Bentler, 1998; Bentler & Yuan, 1999) and therefore a non parametric model estimation procedure was used. EQS (Bentler, 2002) applies a correction to the χ^2 statistic to provide a more accurate solution to the data which is considered non normal. Bentler, (2002) highlights how results may be unreliable if, as is the case in many areas of psychology, the data are non normal but normality assuming statistics are used. In the confirmatory factor analyses of the large WAIS-III standardisation samples this may be less of a problem than in clinical samples where conditions of multivariate normality are less likely to be met.

The group with learning disability were defined as such because they met the criteria of significant or severe impairment in intellectual functioning and were seen in a learning disability service. It was not possible to assess whether the individuals also met the other criteria for a diagnosis of a learning disability – significant impairment in adaptive/social functioning and onset before age of 18 years. An assumption was made that the majority of individuals using the service had a learning disability and only a minority whose intellectual functioning level was higher would be seen if their needs were best met by the learning disability service. Those clearly falling in the borderline IQ range would be more likely to be seen by the non learning disability service following an assessment of their IQ.

Jones *et al.* (2006) chose to exclude individuals with epilepsy, autistic spectrum disorders and those with brain injury on the basis their profiles might confound the results of a more generic low IQ population. In the current study, these exclusion criteria were not applied in order to preserve the authenticity of a clinical population and therefore a difference in results might be expected. The small sample size of the borderline group makes it difficult to draw robust conclusions about differences in cognitive profile. However, it is interesting that a more complex model than that proposed by the test publishers could also account for the data and this may be due to greater inter-subtest scatter resulting from a more heterogeneous data group. Further research on a larger sample would need to be conducted to support this assertion. It would also be of interest to access data from individuals with borderline IQ who are not seen by the learning disability services as these individuals may show a different profile to those seen by psychological services specifically catering to those with a learning disability.

12.3 Administration of WAIS-III to individuals with a learning disability

Research has suggested that clinicians using the WAIS-III with people with a learning disability do not follow standardised procedures (McKenzie *et al.,* 2004). Non standard administration can increase IQ scores and the standard error of measurement (Joncas & Standing, 1998; Slate *et al.,* 1991). Considering these findings, it must be acknowledged that the administration of the WAIS-III to the samples used in this study may have been non standard and as a result affected the scoring of individual subtests and influenced the overall IQ score given to individuals. This is turn may have skewed the outcomes of the analysis.

12.4 Missing data

Some tests can accommodate missing data by using maximum likelihood estimates, assuming the data are randomly missing and the sample represents a normal distribution. The current samples did not meet either of these conditions. The finding that most missing data came from only two subtests is explicable when we look at the subtests required to calculate IQ and index scores. The Symbol Search and Letter Number Sequencing subtests do not need to be taken into account for the IQ scores and as such clinicians may choose to omit these subtests if they do not wish to report index scores. This may be because clinicians recognise the variability in performance on subtests which indicate high levels of differentiation between indexes or it may be due to more practical issues such as time or the perceived likelihood of success on the subtest. In order to calculate index scores and IQ scores the test publishers suggest administering 13 subtests. If only IQ or index scores are required only 11 subtests need to be administered (Tulsky *et al.,* 2003). Research by McKenzie *et al.* (2004) found some clinical psychologists missed out the Letter-Number Sequencing subtest because of clients' difficulty in understanding instructions. As a result of these omissions, it is possible that the data sets used do not accurately reflect those individuals who had only completed 11 subtests. In the initial samples used in this study, 19 per cent of individuals in the LD sample (total n=510) and 22 % in the borderline IQ sample (total n=113) had not completed all 13 subtests.

12.5 Goodness of fit statistics

A comprehensive review of different tests of fit can be found in Bollen and Long (1993) and Arbuckle (2004), however, some of the key limitations are outlined here. In a confirmatory factor analysis the null hypothesis states that the observed and derived correlations are from the same population and any differences in this can be accounted for by sampling error. However, if the null hypothesis is accepted, this does not mean the model is correct, only that it is not wrong (i.e. it is one explanation for the data). In general, a large x^2 suggests the model does not fit the data well and the null hypothesis can be rejected. A smaller x^2 suggests a good model fit. However, with large samples, the x^2 test is too sensitive, leading to the possibility that models with good fit might be rejected (Tabachnick & Fidell, 2001). With small samples, the same statistic may struggle to discriminate two quite different models. It is also a less appropriate test when conditions of normality are not met, and a corrected χ^2 such as the Satorra Bentler χ^2 may be more accurate.

A root mean square of approximation (RMSEA, Browne & Cudeck, 1993) figure of 0.05 or less is often used as an indication of close fit to a model. However, as Arbuckle, (2004) suggests, this figure is subjective and as it takes into account the model's degree of parsimony it can be sensitive to the number of parameters estimated. Researchers can use a number of different fit indices to evaluate their data including the NFI, NNFI, PNFI, ECVI and AIC (Stevens, 2009). The CFI compares the fit of the hypothesised model to a null model which assumes all the variables are unrelated and provides an estimate of fit which takes account of possible misspecification caused by small samples (Bentler, 1990; Schwannauer & Chetwynd, 2007). Like the RMSEA, the CFI can be adjusted using the Satorra-Bentler scaled χ^2 statistic to assess the models which are not multivariate normal.

13. Conclusions

In this study a confirmatory factor analysis of the WAIS-III, weak support was found for the explicit four factor structure in individuals with significant learning disability and those with borderline IQ. For individuals with severe learning disability, fit was poor for all models As well as the four factor models the five factor model consistent with CHC theory provided the best fit to the data from the sample with significant learning disability and a six factor model consistent with an updated version of *gf-gc* theory provided the best fit to the data from those with borderline IQ. These more complex models, consistent with theoretical concepts of crystallised and fluid reasoning, may provide a better explanation of the data when inter-subtest scatter is more heterogeneous. The findings also suggest the cut-off point of IQ 70 is not reflective of an actual difference in cognitive profile as measured by the WAIS-III.

The findings of this research must be viewed in light of the study's limitations not least the smaller sample of people with borderline IQ, the subjective nature of assessing model fit and any bias caused by the likely non standard administration of the WAIS-III to people with a learning disability, and the Flynn Effect. However, the veracity of evaluating model fit in this study has been strengthened by using fit indices appropriate to the sample and employing modifications for data not meeting the conditions for multivariate normality. Although few studies acknowledge the possibility of equivalent models (McCallum *et al.,* 1993), this study has also highlighted more complex models which suggest alternative means of interpreting individual scores when inter-subtest scatter does not follow the pattern suggested by the explicit factor structure of the scale. Future studies may wish to examine further model fit in larger samples with borderline IQ as well as the increased differentiation in verbal scores which were found in this study. It would also be of interest to look in more detail at the pattern of factor loadings and the relative strength of each subtest as a measure of that factor.

The publication of the latest version of the Wechsler adult intelligence scale (WAIS-IV, Psychological Corporation, 2008) highlights the enduring nature of the Wechsler scales. The findings of this study lend some support to the direction taken by the new Wechsler children's and adult scales (WISC-IV & WAIS-IV) in aligning interpretation of the scales more closely to intelligence theory. However, many studies highlight the concerns about using the Wechsler scales with individuals with severe impairment in intellectual ability and this research echoes these concerns. Further research will decide if the latest version addresses adequately some of these issues or if alternative tests are required.

References

Allen, S.R. & Thorndike, R.M. (1995). Stability of the WAIS-R and the WISC-III factor structures using cross validation of covariance structures. *Journal of Clinical Psychology, 51,* 648-657.

American Association on Mental Retardation. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.). Washington, D. C.: American Association on Mental Retardation.

American Association on Mental Retardation (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, D. C.: American Association on Mental Retardation.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: APA.

American Psychiatric Association. (2000). Diagnostic Statistical Manual, Fourth Edition, Text Revised, (DSM-IV-TR), Washington: A.P.A.

Arbuckle, J. L. (2004). *AMOS Users' Guide Version 5.0*. Chicago: Small Waters.

Archer, R.P., Buffington-Vollum, J.K., Stredny, R.V., & Handel, R.W. (2006). A survey of test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 64-94.

Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology, 22,* 1003–1011.

Arnau, R.C. & Thompson, B. (2000). Second-order confirmatory factor analysis of the WAIS-III. *Assessment* , *7* (3), 237-246.

Atkinson, L. & Cyr, J.J. (1988). Low IQ samples and WAIS-R factor structure. *American Journal on Mental Retardation, 93* (3), 278-282.

Aunos, M, Goupil, G. & Feldman, M. (2005) Mothers with Intellectual Disabilities who do or do not have custody of their children. *Journal of Developmental Disabilities, 10* (2), 65-80.

Baddeley, A.D. (1986). *Working memory.* Oxford: Oxford University Press.

Baumeister, A. A. (1987) Mental Retardation: Some conceptions and dilemmas. *American Psychologist*, 42, 796-800.

Bentler, P.M (1990) Comparative fit indexes in structural models. *Psychological Bulletin, 107* (2), 238 -246.

Bentler, P. M. (2007). EQS 6.1. *Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P.M. & Yuan, K-H. (1999). *Structural equation modelling with small samples: Test statistics.* Multivariate Behavioural Research, 34(2), 181-197.

Berry, J.W. (1994). Cross-cultural variations in intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp.316-322). New York: Macmillan.

Berry, J.W. & Irvine, S.H. (1986). Bricolage: Savages do it daily. In R.J. Sternberg & R.K. Wagner (Eds.), *Practical intelligence: nature and origins of competence in the everyday world* (pp.271-306). Cambridge University Press.

British Institute of Learning Disabilities, (2004). Fact-sheet. What is a learning disability? Downloaded, July 2010 from www.bild.org.uk/pdfs/05faqs/ld.pdf

Binet, A. & Simon, T. (1905). Methodes nouvelles por le diagnostic de niveau intellectuelle des anormaux. *L'Annee Psychologique, 11,* 191-244.

Bollen, K.A. & Long, J.S. (1993). *Testing structural equation models.* Newbury Park, CA:Sage.

Boring, E.G. (1923). Intelligence as the tests test it. *New Republic, 35*, 35-37

Braden, J.P. (2003). Accommodating Clients with Disabilities on the WAIS-III and WMS. In D.S. Tulsky.,D.H. Saklofske, G.J. Chelune., R.K. Heaton., R.J. Ivnik., R. Bornstein., A. Prifitera & M. F. Ledbetter. *Clinical Interpretation of the WAIS-III and WMS-III*. London: Academic Press.

British Psychological Society (2002). *Learning Disability: Definitions and Contexts.* Leicester: The British Psychological Society.

Brody, N. (2000). History of Theories and Measurement of Intelligence. In R. Sternberg. *The Handbook of Intelligence.* Cambridge: Cambridge University Press.

Brooks, B.L. & Weaver, L.E. (2005). Concurrent validity of WAIS-III short forms in a geriatric sample with suspected dementia: Verbal, performance and full scale IQ scores. *Archives of Clinical Neuropsychology*, *20*, 1043–1051.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bolen & J.S. Long (Eds.), *Testing structural equation models* (pp.132-162). Newbury Park, CA:Sage.

Burgess, A., Flint, J. & Adshead, H. (1992). Factor structure of the Wechsler adult intelligence scale – revised (WAIS-R): A clinical sample. *British Journal of Clinical Psychology, 31,* 336-338.

Burton, D.B, Ryan, J.J, Axelrod, B.N & Schellenberger, T. (2002). A confirmatory factor analysis of the WAIS-III in a clinical sample with crossvalidation in the standardization sample. *Archives of Clinical Neuropsychology*,*17*, 371-387.

Bryant, F.B & Yarnold, P.R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In Grimm, L.G. and Yarnold, P.R. (Eds.), *Reading and understanding multivariate analysis.* Washington DC: American Psychological Association.

Caine, A., Hatton, C. & Emerson, E. (1998). Service Provision. In E. Emerson, C. Hatton, J. Bromley, & A. Caine. *Clinical Psychology and people with Intellectual Disabilities.* Chichester: Wiley.

Camara, W.J., Nathan. J.S., & Puente, A.E. (2000). Psychological test usage: implications in professional psychology. *Professional Psychology: Research and Practice, 31,* 141-154.

Canivez, G.L. & Watkins, M.W. (2010). *Interpretation of the WAIS-IV with adolescents: Structural validity considerations and concerns.* Paper presented at the 2010 Annual Convention of the National Association of Psychologists, Chicago, IL.

Carroll, J.B. (1978). *The scientific use of factor analysis in the behavioural and life sciences.* New York: Plenum Press.

Carroll, J.B. (1993). Human cognitive abilities. A survey of factor analytic studies. Cambridge: Cambridge University Press.

Carroll, J.B. (1996). A three-stratum theory of intelligence: Spearman’s contribution. In I. Dennis & P. Tapsfield (Eds.), Human abilities: Their nature and measurement. Mahwah, NJ: Lawrence Erlbaum Associates.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Child, D. (2006). The Essentials of Factor Analysis. Third Edition. London: Continuum.

Christensen, B.K., Todd A., Girard, R. & Bagby, M. (2007). Wechsler Adult Intelligence Scale—Third Edition Short Form for Index and IQ Scores in a Psychiatric Population. *Psychological Assessment*, *19*(2), 236-240.

Clausen, H.T. & Scott, R. (1993). Factor analysis of the WAIS-R and Verbal Memory and Visual Memory Indices of the Wechsler Memory Scale--Revised, for a vocational rehabilitation sample. *Perceptual Motor Skills, 76*(3), 907-911.

Clayton, G.A., Sapp, G.L., O' Sullivan, P. & Hall, L. (1986). Comparative Validity of two WAIS-R short forms with vocational rehabilitation clients. *Perceptual and Motor Skills, 63,* 1303-1308.

Cohen, J. (1957). A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology, 21* 451-457.

Comrey, A.L. & Lee, H.B. (1992). *A first course in factor analysis* (2nd Ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Costello, A.B. & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation, 1*(7),1-9.

Crawford, J.R., Allan, K.M., Stephen. D.W., Parker, D.M. & Besson, J.A.O. (1989). The Wechsler Adult Intelligence Scale – Revised (WAIS-R): factor structure in a U.K. sample. *Personality and Individual Differences, 10* (11), 1209-1212.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods,* 1, 16-29.

Davidson, J.E. & Downing, C.L. (2000). Contemporary Models of Intelligence. In R. Sternberg. *Handbook of Intelligence.* Cambridge: Cambridge University Press.

Department of Health. (2001). *Valuing People: A new strategy for Learning Disability in the 21st Century.* Department of Health.

Detterman, D.K., Gabriel, L.T. & Ruthsatz, J.M. (2000). Intelligence and Mental Retardation. In R. Sternberg. *Handbook of Intelligence.* Cambridge: Cambridge University Press.

Dickinson, D., Iannone, V.N. & Gold, J.M. (2002). Factor structure of the Wechsler adult intelligence scale – III in schizophrenia. *Assessment, 9,* (2), 171-180.

Digby, A. (1996). Contexts and Perspectives. In D. Wright & A. Digby (Eds.), *From Idiocy to Mental Deficiency: Historical perspectives on people with learning disabilities.* London:*Routledge.*

Di, Nuovo, S. & Buono, S. (2009). Diagnosis of Intellectual Disability: comparison between clinical criteria and automatized procedures. *Life Span and Disability XII, 2*, 151-159.

Enns, R.A. & Reddon, J.R. (1998). The factor structure of the Wechsler adult intelligence scale- revised: One or two but not three factors. *Journal of Clinical Psychology, 54* (4), 447-459.

Field, A. (2000). *Discovering statistics using SPSS for Windows: advanced techniques for the beginner*. London:Sage.

Fitzgerald, S., Gray, N.S. & Snowden, R.J. (2007). A comparison of WAIS-R and WAIS-III in the lower IQ range: Implications for learning disability diagnosis. *Journal of Applied Research in Intellectual Disabilities, 20*, 323-330.

Flanagan, D. & Kaufman, A. (2004). *Essentials of WISC-IV assessment*. NJ: John Wiley & Sons.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29-51.

Flynn, J.R. (2009). *What is intelligence*? Cambridge: Cambridge University Press.

Ford, K.J., MacCallum, R.C., and Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39,* 219-314.

Frank, G. (1983). The Wechsler enterprise: an assessment of the development, structure, and use of the Wechsler test of intelligence. New York: Pergamon.

Galton, F. (1869). *Hereditary genius: an enquiry into its laws and consequences.* London: McMillan.

Garcia, L.F., Ruiz, M.A. & Abad, F. J. (2003). Factor structure of the Spanish WAIS-III. *Psicothema,15*(1),155-160.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic.

Gardner, H. (1998). Are there additional intelligences? The case for naturalist, spiritual and existential intelligences? In J. Kane (Ed.), *Education, Information and Transformation* (111-132). Englewood Cliffs, N.J.: Prentice Hall.

Gelb, S.A. (1997). The problem of typological thinking in mental retardation. *Mental Retardation, 35,* 448-457.

Gordon, S., Duff, S., Davidson, T. & Whitaker, S. (2010). Comparison of the WAIS-III and WISC-IV in 16-Year-Old Special Education Students. *Journal of Applied Research in Intellectual Disabilities, 23,* 197-200.

Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.) Hillsdale, NJ: Erlbaum.

Greenspan, S. (1999). A contextualist perspective on adaptive behaviour. In R.L. Schalock, and D.L. Braddock (eds.) *Adaptive behaviour and its measurements: Implications for the field of mental retardation* (pp.61-77). Washington, DC: American Association on Mental Retardation.

Greenspan, S. (1999). What is meant by mental retardation? *International Review of Psychiatry, 11*, 6-18.

Greenspan, S. (2006). Mental retardation in the real world: Why the AAMR definition is not there yet. In H.N. Switzky & S. Greenspan (Eds.). What is MR: Ideas for an evolving disability (pp.165-183). Washington, DC: American Association on Mental Retardation.

Gross, R. (1996). *Psychology: The Science of Mind and Behaviour.* Third Edition. London: Hodder & Stoughton.

Groth-Marnat, G. (2000). (Ed). *Neuropsychological Assessment in Clinical Practice*. New York: Wiley.

Groth-Marnat, G. (2009) *Handbook of Psychological Assessment* (5^{th} Edition). Hoboken, NJ: John Wiley & Sons.

Hair, J.F., Anderson, Jr.l., Tatham, R.L. & Black, W.C. (1998). *Multivariate data analysis with readings,(5^{th} ed).*. Englewood Cliffs, NJ: Prentice-Hall.

Harrison, P. L. (1987). Research with adaptive behaviour scales. *The Journal of Special Education, 21*, 37-68.

Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System* (2^{nd} ed.). Minneapolis, MN: Pearson Assessment.

Heijden, P. van der. & Donders, J. (2003). WAIS-III factor index score patterns after traumatic brain injury. *Assessment 10*(2), 115-122.

Hernstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: The Free Press.

Hill, T.D., Reddon, J.R. & Jackson, D.N. (1985). The factor structure of the Wechsler scales: A brief review. *Clinical Psychology Review, 5*, 287-306.

Hishinuma, E.S. (1998). Issues relating to WAIS-R testing modifications for individuals with learning disabilities or attention deficit/hyperactivity disorder. *Learning Disability Quarterly, 21*, 228-240.

Horn, J.L. & Cattell, R.B. (1967). Age differences in fluid and crystallised intelligence. In Gross, R. (1996). *Psychology: The Science of Mind and Behaviour*. Third Edition. London: Hodder & Stoughton.

Horn, J.L. (1986). Intellectual ability concepts. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol.3, pp.35-77). Hillsdale, NJ:Erlbaum.

Jensen, A. R. (2004) cited in McGrew, K.S. (2003). *Cattell-Horn-Carroll CHC (Gf-Gc) Theory: Past, Present and Future.* Downloaded May, 2010 from www.iaspsych.com/CHCPP/B.1st GenerationGf-GcAssessment.

Joncas, J. & Standing, L. (1998). How much do accurate instructions raise scores on a timed test? *Perceptual and Motor Skills, 88,* 1257-1258.

Jones, J, Schaik, P. van. & Witts, P. (2006). The factor structure of the WAIS-III in persons with low IQ. *British Journal of Clinical Psychology, 45* (2), 145-152.

Joreskog, K.G. (1973). General methods for estimating a linear structure equation system. In A.S. Goldberger & O.D. Duncan (Eds.), Structural Equation Models in the Social Sciences (pp.85-112): Seminar Press.

Kaiser, H.F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34*(1), 111-117.

Kamphaus, R.W., Petosky, M.D. & Morgan, A.W. (1997). A history of intelligence test interpretation. 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.

Kanaya, T., Scullin, M. H. & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist*, *58*, 778-790.

Kaufman, A.S. (2000). Tests of intelligence. In R.J. Sternberg (Ed), *Handbook of Intelligence* (pp.445-477). Cambridge: Cambridge University Press.

Kaufman, A.S., Ishikuma, T. & Kaufman-Packer, J.P. (1991) Amazingly short forms of the WAIS-R. *Journal of Psychoeducational Assessment 9*, 4–15.

Kaufman, A.S. & Lichtenberger, E.O. (1999). *Essentials of the WAIS-III Assessment.* New York: Wiley.

Kaufman, A.S. & Lichtenberger, E.O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.) Hoboken, NJ: John Wiley & Sons.

Kaufman, A.S. (2000). Two and three factor solutions of the WAIS-III. *Assessment, 8*(3), 267-280.

Keith, T. (2009). Cited in E.O. Lichtenberger. & Kaufman, A.S. (2009). *Essentials of WAIS-IV Assessment* (pp.30-32). New Jersey: Wiley.

Keith, T.Z., Fine, J.G., Taub, G.E., Reynolds, M.R. & Kranzler, J.H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children- Fourth Edition: What does it measure? *School Psychology Review, 35* (1), 108-127.

Kline, P. (2004). *An easy guide to Factor Analysis*. London: Routledge.

Lawley, D.N. & Maxwell, A.E. (1971). *Factor analysis as a statistical method*. London: Butterworth and Co.

Leckliter, I.N., Matarazzo, J.D. & Silverstein, A.B. (1986). A literature review of the factor analytic studies of the WAIS-R. *Journal of Clinical Psychology, 42,* 332-342.

Lewis, J.R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7* (1), 57-78.

Lichtenberger, E.O. & Kaufman, A.S. (2009). Essentials of WAIS-IV Assessment. New Jersey: Wiley.

Lin, J-D. (2003). Intellectual Disability: Definition, Diagnosis and Classification. *Journal of Medical Science, 23* (2), 83-90.

MacCallum, R.C., Wegener, D.T., Uchino,B.N. & Fabrigar, L.R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

MacCallum, R.C. & Widaman, K. F. (1999). Sample Size in Factor Analysis. *Psychological Methods, 4* (1), 84-99.

McGrew, K.S. (2003). *Cattell-Horn-Carroll CHC (Gf-Gc) Theory: Past, Present and Future.* Downloaded May, 2010 from www.iaspsych.com/CHCPP/B.1st GenerationGf-GcAssessment.

McGrew, K.S. & Flanagan, D.P. (1998). *The intelligence test desk reference.* Boston: Allyn & Bacon.

McKenzie, K., Murray, G.C. & Wright, J. (2004). Adaptations and accommodations: the use of the WAIS-III with people with a learning disability. *Clinical Psychology, 43*, 23-26.

McKenzie, K. & Paxton, D. (2006) Short Forms of Intellectual Assessments & the WASI. In Learning Disabilities, Health, Beliefs and Behaviour. WEBCT. Edinburgh University.

Murray, G., McKenzie, K. & Lindsay, W.R. (2003). A pilot study of the use of the WAIS III with people with a learning disability. *Clinical Psychology,28*, 17-21.

Nihira, K., Leland, H. & Lambert, N. (1993). *AAMR Adaptive Behaviour Scale - Residential and Community* (2nd ed.). Austin Texas: American Association on Mental Retardation.

Nunally, J.O. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Piedmont, R.L., Sokolove, R.L. & Flemming, M.Z. (1992). An evaluation of various WAIS-R factor structures in a psychiatry sample. *Journal of Clinical Psychology, 48* (5), 658-666.

Psychological Corporation, The. (1997). *WAIS-III and WMS-III technical manual*. San Antonio, TX: Author.

Psychological Corporation, The. (2003). *WISC-IV technical and interpretive manual*. San Antonio, TX. Author.

Psychological Corporation, The. (2008) *WAIS-IV Technical and interpretive manual*. San Antonio, TX. Author.

Rabin, L. A., Barr, W.B. & Burton, L.A. (2005) Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS,NAN and APA division 40 members. *Archives of Clinical Psychology, 20*, 33-65.

Reber, A. S. (1985). *Dictionary of psychology*. London: Penguin.

Royal College of Psychiatrists, (2009). Specific learning disabilities, information for parents, carers and anyone working with young people. Downloaded. July 2010 from www.rcpsych.ac.uk/mentalhealthinfo/mentalhealthandgrowingup/specificlearningdifficulties.aspx.

Ryan, J.J., Paulo, A.M., Miller, D.A. & Morrise, J. (1997). Exploratory factor analysis of the Wechsler adult intelligence scale – revised (WAIS-R) in a sample of brain damaged women. *Archives of Clinical Neuropsychology, 12,* (7), 683-689.

Ryan, J.J., Sattler J.M. & Lopez S. J. (2000). Age Effects on Wechsler Adult Intelligence Scale-III Subtests. *Archives of Clinical Neuropsychology,15* (4), 311–317.

Ryan, J.J. & Lopez, S.J. (2001). Wechsler Adult Intelligence Scale-III. In Dorfman & Hersen (Eds.). Understanding Psychological Assessment. New York: Kluwer Academic/Plenum Publishers.

Saklofske, D.H., Hildebrand, D.K. & Gorsuch, R.L. (2000). Replication of the factor structure of the Wechsler adult intelligence scale- third edition with a Canadian sample. *Psychological Assessment, 12* (4), 436-439.

Sanborn, K., Truscott, S. D., Phelps, L. & McDougal, J. L. (2003). Does the Flynn Effect differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment, 21*, 145-159.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514.

Satz, P. & Mogel,S. (1962). An abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology 18*, 77–79.

Schwannauer, M. & Chetwynd, P. (2007). The Brief Symptom Inventory: A validity study in two independent Scottish samples. *Clinical Psychology and Psychotherapy*, *14*, 221-228.

Scottish Executive, (2000). *The same as you? A review of services for people with learning disabilities.* Scottish Executive.

Silverstein, A.B. (1990). Short Forms of Individual Intelligence Tests. Intelligence Tests. *Psychological Assessment, 2,* (1), 3-11.

Slate, J.R., Jones, C.H. & Murray, R.A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale –Revised: An empirical evaluation of practice administrations. (5), 375-379.

Smith, J. (2003). *A factor analysis of the Wechsler Adult Intelligence Scale – 3rd Edition (WAIS-III) with a low IQ population.* Unpublished doctoral thesis, University of Teeside.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behaviour Scales.* Circle Pines, MN: American Guidance Service.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland Adaptive Behavior Scales* (2nd Ed.). Minneapolis, MN: Pearson Assessment.

Spearman, C.E. (1904). *"General Intelligence," objectively determined and measured. American Journal of Psychiatry, 15,* 201-293.

Spearman, C. (1927). *The abilities of man*. London: Macmillan.

Sprague, R.L. & Quay, H.C. (1966). A factor analytic study of the responses of mental retardates on the WAIS. *American Journal of Mental Deficiency, 70* (4), 595-600.

Sternberg, R.J. (1985). Human Intelligence: The model is the message. *Science, 230*, 1111-1118.

Sternberg, R.J. (1997). *Successful intelligence.* New York: Plume.

Sternberg, R. J. (2000). The Concept of Intelligence. In R. Sternberg. *The Handbook of Intelligence*. Cambridge: Cambridge University Press.

Stevens, J. (2009). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, N.J: Lawrence Erlbaum Associates.

Sturmey, P., Gatherer, A., Ghadialli, S., Hallett, S. & Searle, Y. (1993). The Wechsler Adult Intelligence Scale – revised (WAIS-R) factor structure for a British neurologically impaired population. *Personality and Individual Differences, 14* (1), 255-257.

Tabachnick, B.G. & Fidell, L.S. (2001). Using Multivariate Statistics. Fourth Edition. MA: Allyn & Bacon.

Taub, G.E. (2001). A confirmatory factor analysis of the Wechsler Adult Intelligence Scale – third edition: is the verbal/performance discrepancy justified? Practical Assessment, Research & Evaluation, 7(22).

Terman L.M. (1916). The measurement of intelligence: an explanation of, and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale. Boston: Houghton Mifflin.

Thompson, A.P., LoBello, S.G., Atkinson, L., Chisholm, V. & Ryan, J. (2004). Brief Intelligence Testing in Australia, Canada, the United Kingdom and the United States. *Professional Psychology: Research and Practice, 35,* (3), 286-290.

Thurstone, L.L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Truscott, S. D. & Frank, A. J. (2001). Does the Flynn Effect affect IQ scores of students classified as LD? *Journal of School Psychology, 39*, 319-334.

Truscott, S. D., Volker, M. A. (2005). The Flynn Effect and LD Classification: Empirical evidence of IQ score changes that could affect diagnosis. In A. Columbus (Ed.) *Advances in psychology research, Vol. 35* (pp.173-204). New York: Nova Science.

Tulsky, D.S. & Price, L.R. (2003). The Joint WAIS-III and WMS-III Factor Structure: Development and Cross –Validation of a Six-Factor Model of Cognitive Functioning. *Psychological Assessment,15* (2), 149-162.

Tulsky, D.S., Saklofske, D.H. & Zhu, J. (2003). Revising a Standard: An Evaluation of the Origin and Development of the WAIS-III. In D.S. Tulsky., D.H. Saklofske, G.J. Chelune., R.K. Heaton., R.J. Ivnik., R. Bornstein., A. Prifitera & M. F. Ledbetter. *Clinical Interpretation of the WAIS-III and WMS-III*, *Pages 43-92.* London: Academic Press.

Voelker, S. L., Shore, D. L., Brown-More, C., Hill, L., Miller, L. T. & Perry, J. (1990). Validity of self report of adaptive behaviour scales by adults with mental retardation. *Mental Retardation,28* (5), 305-309.

Ward, (1990). Prediction of verbal, performance, and full scale IQ's from seven subtests of the WAIS-R. *Journal of Clinical Psychology 46*, 436–440.

Wechsler, D. (1944). The measurement of adult intelligence. In Gross, R. (1996). *Psychology: The Science of Mind and Behaviour.* Third Edition. Hodder & Stoughton: London.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale (WAIS)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised (WAIS-R).* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). Manual *for the Wechsler Intelligence Scale for Children- Third Edition (WISC-III),* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale - Third Edition. London: Psychological Corporation.

Whitaker, S. (2005). The uses of the WISC-III and the WAIS-III with people with a learning disability: Three concerns. *Clinical Psychology Forum, 50*, 37-40.

Whitaker, S. (2008). WISC-IV and low IQ: review and comparison with the WAIS-III. *Educational Psychology in Practice, 24* (2), 129-137.

Whitaker, S. (2010a). Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III. *Personality and Individual Differences*, 48, 517–521

Whitaker, S. (2010b). Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III. *Personality and Individual Differences*, 48, 517–521

Wymer, J.H., Rayls, K. & Wagner, M.T. (2003). Utility of a clinically derived abbreviated form of the WAIS-III. *Archives of Clinical Neuropsychology 18* (8) 917-927.

Yao, S., Chen, H, Jiang, L. & Wai-Cheong, C.T. (2007). Replication of factor structure of Wechsler Adult Intelligence Scale – III Chinese version in Chinese mainland non-clinical and schizophrenia samples. *Psychiatry and Clinical Neurosciences, 61,* 379-384.

Yerkes, R.M. (Ed.). (1921). *Psychological examining in the United States Army.* Memoirs, National Academy of Sciences, Vol.15.

Yuan, K–H. & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 51*, 289-309.

Appendix 1: Description of WAIS-III subtests (modified from Ryan & Lopez, 2001, p.22).

Subtest	**Brief Description**
Vocabulary	Examinees are required to define words, presented in order of increasing difficulty. A measure of word knowledge.
Similarities	Examinees are required to identify similarities between pairs of words. A measure of verbal concept formation and abstract thinking.
Arithmetic	A timed subtest involving counting, addition, multiplication, subtraction and basic probability.
Digit Span	Examinees are required to repeat increasing numbers of digits forward (2-9 digits). It appears to be a measure of mental tracking and auditory sequencing. Digits backward appears to measure short term memory, mental tracking and internal visual scanning.
Information	Requires examinees to answer general knowledge questions about objects, events and places.
Comprehension	Requires examinees to answer questions about everyday situations using their knowledge of social convention and common sense.
Letter-Number Sequencing (supplementary subtest which can replace digit span)	Examinees are required to order random strings of numbers and letters presented orally. It appears to measure mental flexibility, divided attention and auditory tracking.
Picture Completion	A timed subtestsin which examinees are required to identify the missing part from drawings on people, objects, scenes and animals. It appears to measure the ability to discriminate between essential and non essential details.
Digit-Symbol Coding	A timed task requiring examinees to draw a symbol in a lower part of a box when the upper part contains a number associated with a symbol in the key. The subtest requires visual tracking, paired-associate learning and visual scanning.

Block design	Examinees are timed to reproduce a design using coloured blocks. It appears to measure visuo-spatial problem solving, constructional ability and non verbal concept formation.
Matrix Reasoning	Requires the examinees to choose the correct answer on non verbal items involving serial reasoning, classification and pattern completion. It is said to be a measure of fluid intelligence.
Picture Arrangement	A timed subtest required examinees to place a series of pictures in a logical sequence. It appears to measure visual sequencing and planning ability within a social context.
Symbol Search (A supplementary subtest which can be substituted for digit symbol coding)	A timed subtest requiring examinees to identify if a symbol presented on the left is also presented on the right.
Object Assembly (An optional subtest which can be substituted for any performace scale subtest provided the examinee is in the 16-74 age range.	A timed subtest requiring examinees to solve jigsaw puzzles. It seems to measure visual organisation, constructional ability as well as an understanding of relationships in part and as a whole.

Appendix 2: Approval from the Caldicott Guardian

Lothian NHS Board

Public Health & Health Policy
Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9000
Fax 0131 536 9009
www.nhslothian.scot.nhs.uk

NHS
Lothian

Dr Hannah MacLean
Trainee Clinical Psychologist
15/1 The Steils
EDINBURGH
EH10 5XD

Date 20th November 2008
Your Ref
Our Ref JS/EM/0878

Enquiries to Jim Sherval
Extension 89374
Direct Line 0131 536 9374
Direct Fax 0131 536 9164
Email Jim.Sherval@nhslothian.scot.nhs.uk

Dear Dr MacLean

Caldicott Approval: Factor analysis of the Wechsler Adult Intelligence Scale 3rd Edition in sample with borderline and low IQ and with a learning disability sample

Thank you for the information supplied for this study.

Request received from	Dr Hannah MacLean, Trainee Clinical Psychologist
Summary of proposal	Data is to be collected by a trainee clinical psychologist employed by NHS Lothian for a thesis project required for completion of the Doctorate in Clinical Psychology at Edinburgh University. The data required is information that is routinely collected by clinical psychologists working in the field of learning disability. The data will comprise numerical scores on a psychological test in addition to basic descriptive information about age and gender. The data will be collected by hand from existing case notes and then processed by a statistical method called a factor analysis. The information will already have been gathered as part of routine assessment and therefore no contact with patients will be required.
Patient identifiable information requested	Age, Gender
Approved	Yes
Reason for decision	Benefit to patients via higher level training

I look forward to seeing the results of this study.

Yours sincerely

Dr Alison McCallum
Director of Public Health & Health Policy

cc Ann Buchan, Business Manager, Public Health & Health Policy
Professor Heather Cubie, Research & Development Director, RIE

Appendix 3: Data Recording Sheet

M/ F	Age	FS IQ	VO	Si	In	Co	PC	BD	MR	PA	Ar	DS	LN	DC	SS	ID

Appendix 4: Clincian's Recording Sheet

(To be destroyed when clinician is notified by the researcher. Until then, please keep confidential)

Name	ID code

Appendix 5 Standardised solutions and squared multiple correlations for remaining models

Table 26. Standardised solutions and squared multiple correlations for one factor model (sample of people with severe learning disability)

Subtest	Loading on g	Squared multiple correlations
PC	.117	.014
VO	.341	.117
DC	.602	.363
SI	.139	.019
BD	.599	.359
AR	.477	.227
MR	.451	.203
DS	.486	.237
IN	.269	.072
PA	.305	.093
CO	.264	.070
SS	.421	.178
LN	.466	.217

Table 27. Standardised solutions and squared multiple correlations for a two factor model (sample of people with severe learning disability)

Subtest	Loading on Verbal Factor	Loading on Performance Factor	Squared multiple correlation
PC		.187	.035
VO	.412		.169
DC		.654	.427
SI	.130		.017
BD		.612	.375
AR	.627		.393
MR		.473	.223
DS	.620		.385
IN	.229		.052
PA		.350	.122
CO	.275		.076
SS		.517	.267
LN	.473		.224

Table 28. Loadings on second order factor and squared multiple correlations for a two factor model (sample of people with severe learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Verbal	.588	.346
Performance	1.00	1.00

Table 29. Standardised solutions and squared multiple correlations for a three factor model (sample of people with severe learning disability)

Subtest	Loadings on Verbal Factor	Loadings on Performance Factor	Loadings on Attention Factor	Multiple Squared Correlation
PC		.186		.035
VO	.672			.452
DC			.668	.477
SI	.393			.155
BD		.714		.510
AR	.375			.141
MR		.477		.228
DS			.448	.201
IN	.257			.066
PA		.313		.098
CO	.491			.241
SS			.432	.187
LN			.445	.198

Table 30. Loadings on second order factor and squared multiple correlations for a three factor model (sample of people with a severe learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Verbal	.478	.228
Performance	.909	.827
Attention	.943	.888

Table 31. Standardised solutions and squared multiple correlations for a four factor factor model (13 subtests; sample of people with severe learning disability)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Squared Multiple Correlation
PC		.177			.031
VO	.644				.414
DC				.761	.578
SI	.455				.207
BD		.681			.463
AR			.632		.399
MR		.500			.250
DS			.670		.448
IN	.282				.080
PA		.333			.111
CO	.525				.276
SS				.523	.274
LN			.506		.256

Table 32. Loadings on second order factor and squared multiple correlations for a four factor model (13 subtests; sample of people with severe learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
VCI	.408	.166
POI	.871	.758
WMI	.644	.414
PSI	.838	.702

Table 33. Standardised solutions and squared multiple correlations for a six factor model (sample of people with severe learning disability)

Subtest	Loadings on Semantic Memory Factor	Loadings on Verbal Reasoning Factor	Loadings on Constructional Praxis	Loadings on Visual Reasoning Factor	Loadings on Working Memory Factor	Loadings of Processing Speed Factor	Squared Multiple Correlation
PC			.243				.059
VO	.347						.120
DC						.792	.628
SI		.392					.108
BD			1.000				1.000
AR		-.065			.646		.391
MR				.735			.541
DS					.668		.446
IN	.308						.095
PA				.418			.175
CO		.503					.253
SS						.502	.252
LN					.515		.265

Table 34. Loadings on second order factor and squared multiple correlations for a six factor model (sample of people with significant learning disability)

Factor	**Loading on 2nd order factor**	**Squared Multiple Correlation**
Semantic Memory	1.000	1.000
Verbal Reasoning	.538	.289
Constructional Praxis	.633	.401
Visual Reasoning	.617	.381
Working Memory	.676	.457
Processing Speed	.729	.532

Table 35. Standardised solutions and squared multiple correlations for one factor model (sample of people with significant learning disability)

Subtest	**Loading on g**	**Squared multiple correlations**
PC	.411	.169
VO	.192	.037
DC	.521	.271
SI	.144	.021
BD	.494	.244
AR	.306	.094
MR	.353	.124
DS	.141	.020
IN	.206	.043
PA	.510	.261
CO	.104	.011
SS	.670	.448
LN	.414	.171

Table 36. Loadings on second order factor and squared multiple correlations for a two factor model (sample of people with significant learning disability)

Subtest	Loading on Verbal Factor	Loading on Performance Factor	Squared multiple correlation
PC		.459	.211
VO	.700		.490
DC		.527	.278
SI	.537		.288
BD		.493	.243
AR	.160		.026
MR		.350	.122
DS	.288		.083
IN	.655		.428
PA		.526	.276
CO	.479		.230
SS		.719	.516
LN	.361		.130

Table 37. Standardised solutions and squared multiple correlations for a two factor model (sample of people with significant learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Verbal	.361	.130
Performance	.414	.171

Table 38. Standardised solutions and squared multiple correlations for a three factor model (sample of people with significant learning disability)

Subtest	Loadings on Verbal Factor	Loadings on Performance Factor	Loadings on Attention Factor	Squared Multiple Correlation
PC		.524		.274
VO	.685			.469
DC			.575	.331
SI	.546			.298
BD		.527		.277
AR	.100			.010
MR		.354		.125
DS			.069	.005
IN	.670			.449
PA		.592		.350
CO	.519			.269
SS			.783	.613
LN			.340	.116

Table 39. Loadings on second order factor and squared multiple correlations for a three factor model (sample of people with a significant learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Verbal	.148	.022
Performance	.849	.721
Attention	.874	.764

Table 40. Standardised solutions and squared multiple correlations for a four factor factor model (13 subtests; sample of people with significant learning disability)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Squared Multiple Correlation
PC		.510			.260
VO	.682				.465
DC				.555	.308
SI	.553				.306
BD		.553			.284
AR			.441		.194
MR		.361			.130
DS			.464		.215
IN	.668				.446
PA		.591			.349
CO	.519				.269
SS				.827	.684
LN			.642		.412

Table 41. Loadings on second order factor and squared multiple correlations for a four factor model (13 subtests; sample of people with significant learning disability)

Factor	Loading on 2^{nd} order factor	Squared Multiple Correlation
VCI	.147	.022
POI	.826	.682
WMI	.445	.198
PSI	.866	.750

.

Table 42. Standardised solutions and squared multiple correlations for a six factor model (sample of people with significant learning disability)

Subtest	Loadings on Semantic Memory Factor	Loadings on Verbal Reasoning Factor	Loadings on Constructional Praxis	Loadings on Visual Reasoning Factor	Loadings on Working Memory Factor	Loadings of Processing Speed Factor	Squared Multiple Correlation
PC			.484				.424
VO	.725						.540
DC						.549	.632
SI		1.00					.557
BD			.532				.611
AR		-.132			.491		.479
MR				.364			.280
DS					.478		.475
IN	.655						.582
PA				.579			.538
CO		.347					.445
SS						.836	.589
LN					.617		.474

Table 43. Loadings on second order factor and squared multiple correlations for a six factor model (sample of people with significant learning disability)

Factor	Loading on 2nd order factor	Squared Multiple Correlation
Semantic Memory	.192	.037
Verbal Reasoning	.113	.013
Constructional Praxis	.989	.978
Visual Reasoning	1.000	1.000
Working Memory	.433	.188
Processing Speed	.739	.546

Table 44. Standardised solutions and squared multiple correlations for one factor model (sample of people with borderline IQ)

Subtest	Loading on g	Squared multiple correlations
PC	.346	.120
VO	-.965	.931
DC	.172	.030
SI	-.435	.189
BD	.281	.079
AR	.213	.045
MR	.206	.042
DS	.042	.002
IN	-.392	.154
PA	.289	.084
CO	-.609	.371
SS	.239	.057
LN	-.099	.010

Table 45. Standardised solutions and squared multiple correlations for a two factor model (sample of people with borderline IQ)

Subtest	Loading on Verbal Factor	Loading on Performance Factor	Squared Multiple Correlation
PC		.507	.257
VO	1.0		1.00
DC		.295	.087
SI	.421		.177
BD		.385	.148
AR	-.212		.045
MR		.355	.126
DS	-.040		.002
IN	.369		.136
PA		.564	.318
CO	.590		.348
SS		.458	.210
LN	.105		.011

Table 46. Standardised solutions and squared multiple correlations for a three factor model (sample of people with borderline IQ)

Subtest	Loadings on Verbal Factor	Loadings on Performance Factor	Loadings on Attention Factor	Squared Multiple Correlation
PC		.530		.281
VO	.984			.968
DC			.544	.296
SI	.429			.184
BD		.411		.169
AR	-.216			.047
MR		.389		.152
DS			.180	.032
IN	.380			.145
PA		.609		.370
CO	.601			.361
SS			1.0	1.000
LN			.289	.084

Table 47. Standardised solutions and squared multiple correlations for a four factor model (11 subtests; sample of people with borderline IQ)

Subtest	Loadings on VCI Factor	Loadings on POI Factor	Loadings on WMI	Loadings on PSI	Squared Multiple Correlation
PC		.493			.243
VO	.998				.997
DC				.544	.296
SI	.422				.178
BD		.344			.118
AR			.327		.107
MR		.163			.027
DS			.778		.605
IN	.370				.137
SS				1.000	1.000
LN			.572		.327

Table 48. Standardised solutions and squared multiple correlations for a five factor model (sample of people with borderline IQ)

Subtest	Loadings on Gc Factor	Loadings on Gv Factor	Loadings on Gf	Loadings on Gsm	Loadings on Gs	Squared Multiple Correlation
PC		.386				.149
VO	.983					.967
DC					.546	.298
SI	.557					.310
BD		.305				.093
AR			.000	.380		.138
MR			.000			.000
DS				.740		.548
IN	.353					.125
CO	.594					.353
SS					1.000	1.000
LN				.597		.357