Intelligence 38 (2010) 481–496

Contents lists available at ScienceDirect

Intelligence

Confirmatory factor analysis of Project Spectrum activities. A second-order g factor or multiple intelligences?

Juan L. Castejon *, Antonio M. Perez, Raquel Gilar

Department of Developmental Psychology and Didactics, University of Alicante, Spain

A R T I C L E I N F O

Article history:
Received 23 October 2007
Received in revised form 3 May 2010
Accepted 8 July 2010
Available online 3 August 2010

Keywords:
Multiple intelligences
Confirmatory factor analysis
g factor

A B S T R A C T

This paper compares different theoretical models of the structure of intelligence, based on the analysis of data obtained in a series of measured abilities corresponding to the Spectrum assessment activities (Gardner, Feldman & Krechevsky, 1998) in a sample of 393 children enrolled in kindergarten and first grade. The data were analyzed using confirmatory factor analysis. The models compared were: a) a model with six first-order uncorrelated factors, b) a model with six first-order factors and one second-order general factor, *g*; c) a model with two correlated second-order general factors, in which the cognitive intelligences load on a "cognitive" general factor and the non-cognitive intelligences load on a "non-cognitive" general factor, and d) a model with six first-order correlated factors. The percentage of variance in measured abilities due to *g* and to first-order factors was also estimated. Overall, the results indicate that the Spectrum activities are not as separate from *g* as proposed by the defenders of multiple intelligences theory, nor as unitary as argued by the defenders of *g* factor models.

Psychometric and differential traditions in research into intelligence have generated a very broad set of research results regarding mental abilities and their structure. Most studies are based on correlational methods, which chiefly use exploratory and confirmatory factor analysis techniques (Brody, 2000).

One of the main goals of this approach to the study of intelligence is to identify the number of distinguishable factors or aptitudes that exist, as well as to establish the possible structure of relationships between these mental abilities. The results of a wide range of research projects reveal the existence of a large group of factors (Carroll, 1993): the verbal factor, containing verbal material; the spatial visualization factor; numerical reasoning; mechanical reasoning; and the memory factor, referring to recall of specific previously acquired information.

According to the *g* factor theory, there is also one large general factor over and above all of these group factors, which encompasses the common variance between the above mentioned factors. This factor becomes clearer when a diverse set of cognitive tasks and a larger more representative sample of the general population are considered (Carroll, 1993; Jensen, 1998). Its existence was originally hypothesized by Spearman (1904), who labeled it simply *g*.

The crystallization of an empirically-based psychometric taxonomy of mental abilities occurred in the late 1980s to early 1990s (McGrew, 2005). During the past decade the Cattell–Horn *Gf-Gc* and Carroll, CHC, three-stratum models have emerged as the consensus psychometric-based theory for understanding the structure of human intelligence and as a working taxonomy to test and evaluate structural models of human intelligence (McGrew, 2009). For example, Johnson and Bouchard (2005) and Johnson, Nijenhuis, and Bouchard (2008) applied confirmatory factor analysis (CFA) methods to datasets analyzed by Carroll. They used CFA methods to compare versions of the Carroll, Cattell–Horn *Gf-Gc*, Vernon verbal–perceptual model, and Johnson and Bouchard verbal–perceptual–rotation (VPR) model. Support for the VPR model

* Corresponding author. Departamento de Psicología Evolutiva y Didáctica, Universidad de Alicante, Carretera San Vicente s/n, 03080 Alicante, Spain.
E-mail address: jl.castejon@ua.es (J.L. Castejon).

0160-2896/$ – see front matter
doi:10.1016/j.intell.2010.07.002

was presented via the CFA analyses. This constitutes a refinement and extension of the CHC taxonomy.

In contrast with the *g* factor theory, a number of scholars defend positions that challenge the strong version of IQ that emerged from the psychometric tradition (Gardner, 2003, 2006; Horn & Cattell, 1966). In these theories, intelligence is seen as having several, at least partially, dissociable aspects, and the primacy of *g* is questioned. The term intelligence refers not only to the general factor but also to several broad organizations of abilities and more narrow primary specific factors (Carroll, 1993; Horn & Noll, 1994). Intelligence is the full hierarchical structure of abilities as conceived by these authors, not just the highest-order factor, *g*. The existence of a single higher-order general factor *g* is the focus of much debate, even among the supporters of the CHC theory (Horn, 2007; McGrew, 2005).

Drawing on evidence from a range of disciplines, including biology, anthropology, and psychology, Gardner (1983, 1999) concluded that humans have a number of relatively autonomous intellectual capacities, called *multiple intelligences*. Gardner's theory diverges from certain traditional conceptions. Like other theorists (Ceci, 1990/1996; Sternberg, 1985, 2003; Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001; Thurstone, 1938), Gardner argued for a notion of intelligence that included non-cognitive abilities as opposed to other theories such as those proposed by Jensen (1998), and Carroll (1993).

Gardner (1983/1993) defined intelligence as the ability to solve problems or to fashion new products that are valued in at least one culture. The major claim in the theory is that the human intellect is better described as consisting of a set of semi-autonomous computational devices, each of which has evolved to process certain kinds of information in certain kinds of ways. Each of the major intelligences is itself composed of subintelligences. To what extent these subcomponents correlate with one another is an empirical question (Gardner, 1993, 2006).

Gardner (1983, 2006) argued that standard intelligence tests typically probe only *linguistic* and *logical–mathematical* intelligences, and certain forms of *spatial* intelligence. In Gardner's view, there are at least five other human intelligences: *musical* intelligence, *bodily-kinesthetic* intelligence, *naturalistic* intelligence, *interpersonal* intelligence, and *intrapersonal* intelligence. According to Gardner, all human beings possess all of the intelligences, but we differ in relative strengths and weaknesses. Each of these intelligences is concisely and fully described in Gardner (1999, pp. 41–43).

The degree of correlation among intelligences is another open question in Gardner's theory: "Nowadays an increasing number of researchers believe the opposite; that there exists a multitude of intelligences, *quite independent* [italics added] of each other; that each intelligence has its own strengths and constraints;" (Gardner, 1993, p. xxiii). This corresponds to an initial or strong version of multiple intelligences theory.

However, in more recent developments Gardner recognized that: "The degree of correlation among intelligences is yet to be determined (because we do not have adequate measures for most of the intelligences). In any event, there is no reason to think that they will be dominated by a single *g* factor"(Gardner, 2003, p. 47). Nor did Gardner agree that the multiple intelligences may be perceived as "special talents" within this general factor (Gardner, 2006). This view that permits intelligences to correlate can be defined as the recent or weak version of multiple intelligences theory.

Although some critics (Brody, 2006; Visser, Ashton, & Vernon, 2006a) claim that there is no empirical evidence to support a theory of multiple intelligences, Gardner (1983) examined several empirical studies when identifying the seven intelligences. Nevertheless, only a few correlational studies exist that support Gardner's theory, most studies are experimental and based on clinical evidence. The lack of correlational studies providing empirical support for Gardner's theory of multiple intelligences is due to several reasons, including the argument of the theory itself against using standardized tests to measure intelligence, and the lack of appropriate tools to do so, as Gardner (2003, 2006) himself admits.

Only a few studies have tested the structural validity of this theory using correlational methodology, and exploratory and confirmatory factor analysis techniques. The aim of these studies was to confirm the presence of different types of intelligence in a battery of activities derived from Project Spectrum. Plucker, Callahan, and Tomchin (1996) performed exploratory factor analysis in order to test the existence of four types of intelligence – spatial, logical/mathematical, linguistic and interpersonal – in a sample of 1813 children in kindergarten and first grade, using the Multiple Intelligences Assessment Technique, which is based upon the assessment activities used in Project Spectrum (Gardner et al., 1998). The technique consisted of 13 performance-based activities, teacher ratings, and observational checklists corresponding to the four intelligences. The factor analysis – principal component extraction and varimax rotation – supported the presence of the linguistic, logical–mathematical and spatial subscales, and the combination of interpersonal and linguistic intelligence activities in the first factor. Although these factor analysis results appear to provide some support for the theory of multiple intelligences, they are limited by the fact that they were obtained using exploratory factor analysis, rather than CFA, a much better procedure to study this issue.

Pyryt (2000) reanalyzed the correlation matrix of Plucker et al. (1996) to illustrate how higher-order exploratory factor analysis using more adequate procedures – maximum likelihood and direct oblimin – might be used to explain the constructs found in the initial factor analysis. Consistent with Carroll's (1993) factor analysis study of mental abilities, results indicated that the *g* factor underlies correlations between first-order factors.

Gridley (2002) reanalyzed data from Plucker et al. (1996) to illustrate how the use of CFA might help to determine the factorial structure that fits these empirical data. The findings obtained by Gridley (2002) showed that a model with several correlated factors fitted the data from Plucker et al. (1996) better than did a hierarchical model with *g* at the top.

The results obtained by Gridley (2002), using a higher-order CFA model, showed that, although as might be expected some tasks were more highly *g*-loaded or *g*-influenced than others, each individual task retained variance that was not attributable to *g*, which suggests that the individual tasks do tap into abilities other than *g*.

Visser, Ashton, and Vernon (2006a) investigated Gardner's Theory of Multiple Intelligences by examining the relationships between the eight hypothesized intelligences. In particular, they made a distinction between two categories of intelligences in Gardner's model, according to the extent to which cognitive (linguistic, spatial, logical/mathematical, naturalistic, and interpersonal) and non-cognitive (intrapersonal, bodily-kinesthetic, and musical) abilities were involved in each domain. The findings support the hypothesis that all of the tests except those in the partly non-cognitive domains are correlated with each other and with an independent, external measure of the *g* factor (Visser et al., 2006a). They also predicted that a factor analysis of the tests would yield a *g* factor, with all of the purely cognitive tests having substantial *g*-loading, and with all of the remaining tests having lower *g*-loading. Furthermore, Visser et al. (2006a) found a substantial correlation within each domain, beyond the influence of *g*. The residual correlations between the tests within each of the logical–mathematical, spatial and linguistic domains were large enough to suggest a considerable influence of non-*g* sources of variance on the relationships between the tests. This is consistent with the theory of multiple intelligences. However, this result is similarly explained by hierarchical models of intelligence which postulate several group factors in addition to the *g* factor.

The main goal of this study was to contrast different theoretical models: a) a model with six first-order uncorrelated factors, corresponding to a strong version of the theoretical model of multiple intelligences – Model 1- (Gardner, 1983/Gardner, 1993); b) a model with six first-order factors and one second-order general factor, the general intelligence factor *g* (Carroll, 1993) – Model 2 -; c) a model with two correlated second-order general factors, in which the cognitive intelligences (linguistic, spatial, logical/mathematical, and naturalistic) load on a “cognitive” general factor *g*, and the non-cognitive intelligences (bodily-kinesthetic and musical) load on a “non-cognitive” general factor (Visser et al., 2006a) – Model 3 -; and d) a model with six first-order correlated factors, corresponding to a weaker but more recent version of the theory of multiple intelligences (Gardner, 2006) – Model 4 -.

These theoretical models were compared using confirmatory factor analysis, which allows for rigorous empirical testing of competing theoretical models of human cognitive structure of abilities (Gridley, 2002; Horn, 2007; McGrew, 2005). The use of CFA is especially recommended for the analysis of tests that are based on strong underlying theory (Reynolds & Keith, 2006).

1. Method

1.1. Participants

The initial sample consisted of 411 children, 18 (4.37%) of whom were excluded from the study because they had special educational needs. The final sample for this study consisted of 393 children enrolled in kindergarten (n = 186) and first grade (n = 207), in two private and three state primary schools in two large cities in Spain, with a student population drawn from urban and suburban areas. Because of the racial and ethnic homogeneity of the country, the majority of children were Caucasian (99%). The sample was obtained intentionally; that is, the private and state schools were chosen to represent the school population. All the children in each class – kindergarten and first grade– in each school took part in the study. All the parents gave their informed consent for their children to participate. The sample was made up of all pupils enrolled in 16 class groups. Childhood socio-economic status (SES) was indexed according to parental occupation. There was a wide range of socio-economic statuses with a predominance of middle class children.

Girl pupils accounted for 50.6% (n = 199) of the sample. Ages ranged from 4 to 7 years (M = 5.34; SD = 1.76). The sample population studied was representative of the national general population (girls = 50.05%; boys = 49.95%, and children in kindergarten = 50.69%; children in first grade = 49.31%). The differences in percentages between sample and population were not statistically significant (higher CR = 1.33 < 1.96).

1.2. Instruments and measures

One of the problems with the theory of multiple intelligences is the difficulty in finding valid measurement instruments for the different intelligences that enable the theory to be tested. Gardner provided little guidance as to how his intelligences might be tested (Visser, Ashton, & Vernon, 2006b) and Visser et al. (2006a) encouraged Gardner to provide measures for his eight intelligences so that his theory could be put to the test.

As Gardner (2006) recognized “except for Project Spectrum (Gardner, Feldman, & Krechevsky, 1998), I have not devoted energies to the devising of tasks that purport to assess MI” (p. 504).

In order to evaluate multiple intelligences (MI), 8 activities were used, which were designed by Gardner et al. (1998) as part of Project Spectrum with a view to evaluating six of the intelligences proposed by Gardner. The purpose of the 8 activities used here was to evaluate the abilities of the study participants in relation to the different intelligences considered in our model.

The technique consists of 8 performance-based activities corresponding to six of the multiple intelligences. These activities were drawn from Project Spectrum at Harvard University (Gardner et al., 1998).

The assessment includes 22 abilities (related to 8 activities) – each made up of a specific number of measured items as explained below in the description of each ability – evaluating six different intelligences: *naturalistic, bodily-kinesthetic, spatial, musical, logical–mathematical, and linguistic* intelligence. The assessment and scoring criteria are based on Gardner et al. (1998) with local modifications. The activities were adapted and contextualized by experts in the research team.

For each item included in each ability, pupils' performance was rated on a scale as “never” (1) — not evident or not observed during performance of the activity; the evaluated ability was not demonstrated during the activity in question; “sometimes” (2) — sometimes evident or sometimes observed during the activity; “almost always” (3) — almost always evident or almost always observed during the activity; or “always” (4) — always evident or always observed during the activity.

Table 1 shows a summary of the intelligences evaluated, the activities used, the abilities measured in each intelligence, and the specific items that are rated in each of the abilities.[1]

Below is a list of the intelligences assessed and the activities used:

1. *Naturalistic intelligence.* The activities used to assess the scientific domain were: a) Discovery area (naturalistic measure). The discovery area is a permanent area of the classroom devoted to natural science activities. Activities include caring for small animals, growing plants and examining a range of natural materials such as rocks and shells. b) Sink and float activity (hypothesis-testing measure). The sink and float activity is used to assess a child's ability to generate hypotheses based on their observations and to conduct simple experiments. The child is shown a tub of water and an assortment of floating and sinking materials. Then, he/she is asked to make a series of predictions about the objects and to generate a hypothesis to explain their behavior. The child is also encouraged to try out their own ideas for exploring and experimenting with the materials. The abilities evaluated were: identification of relationships (n1), close observation (n2), hypothesis formation (n3), experimentation (n4), interest in nature activities (n5), and knowledge of the natural world (n6). Taking for example the item of the ability n1 "notices similarities and dissimilarities between objects", the child was shown a picture of a dog and a cat and pointed out similarities and differences. The rating system was as follows: 1 point = does not point out any similarities or differences; 2 points = uses up to 2 criteria; 3 points = uses up to 4 criteria; and 4 points = uses more than 4 criteria.
2. *Bodily-kinesthetic intelligence.* The activity used to assess bodily-kinesthetic intelligence was creative movement, in which children participate in sessions that include activities like Simon Says; moving different parts of the body at the same time; and responding to music, props, and verbal images. Every session ends with free dance to music. The abilities evaluated were: sensitivity to rhythm (bk1), expressiveness (bk2), body control (bk3), and generation of movement ideas (bk4). Taking for example the item of the ability bk2 " responds vividly to different verbal images", the child was asked to act out situations described by the rater (you are as stiff as a robot, floating like a soap bubble, zigzagging like a snake, made of rubber, skating on ice, moving like a clockwork toy and then you run down) and rated as follows: 1 point = is not able to act out any of the situations or acts out only 1 situation correctly; 2 points = acts out at least 2 situations correctly; 3 points = acts out at least 4 situations correctly; and 4 points = acts out 6 situations correctly.
3. *Spatial intelligence.* The art portfolio activity was used to assess this intelligence. Throughout the school year, each child's artwork was collected in a portfolio. These portfolios included drawings, paintings, collages, and three-dimensional pieces. The abilities evaluated were: level of representation (s1), degree of exploration (s2), and level of artistry (s3). Taking for example the item of the ability s1 "spatial integration" the child was rated as follows: 1 point = the figures/objects seem to float in space; 2 points = the figures/objects do not float but are placed on the ground; 3 points = the figures/objects are placed correctly on the ground; and 4 points = the figures/objects are placed correctly on the ground and bear relation to each other and to the drawing as a whole.
4. *Musical intelligence.* The singing activity was used to assess musical intelligence. The singing activity is designed to assess the child's ability to maintain accurate pitch and rhythm while singing, and their ability to recall a song's musical properties. The abilities evaluated were: rhythm (m1), pitch (m2), and general musical ability (m3). Taking the item of the ability m2 "interval: consistently maintains the correct interval between notes", the rating system was: 1 point = never; 2 points = sometimes; 3 points = almost always; and 4 points = always.
5. *Logical-mathematical intelligence.* This intelligence was assessed using the dinosaur game activity. This activity is designed to measure the child's understanding of numerical concepts, counting skills, ability to adhere to rules and use of strategy. The abilities evaluated were: numerical reasoning (lm1), spatial reasoning (lm2), and logical reasoning (lm3). Taking the item of the ability lm3 "uses the symbols correctly: when the child rolls the dice and gets a + he/she moves forward and when he/she gets a − he/she moves backward", the rating system was: 1 point = never; 2 points = sometimes; 3 points = almost always; and 4 points = always.
6. *Linguistic intelligence.* The language activities were: a) Storyboard activity. This activity is designed to provide a concrete but open-ended framework in which a child can create stories. Children are asked to tell a story using storyboard equipment with an ambiguous-looking landscape, foliage, dwellings, and assorted figures, creatures, and props. b) Reporter activity. This activity assesses children's ability to describe an event they have experienced. Children are asked to tell the rater what happened over the weekend. They may relate personal interactions or focus on describing events and activities. The abilities evaluated were: primary language functions (l1), narration ability (l2), and information ability (l3). As an example, the item of the ability l2 "nature of narrative structure" was rated as follows: 1 point = not evident or not observed; 2 points = the story only describes the main event; 3 points = the child gives names and/or roles to the characters. He mentions their relationships but does not elaborate on them; and 4 points = the child invents, follows the plot, identifies one or more characters and creates relationships between them, together with their physical and emotional characteristics and other details.

1.3. Procedure

Interviews were held with the school principals to describe the aims of the research project and the proposed assessment procedure and to obtain their permission and collaboration. The teachers were selected on the basis of their willingness to participate in the research. Subsequently, a

[1] For further information about the items evaluated and the specific criteria used to rate each of the items, please visit: http://www.ua.es/dpto/dped/rating%20criteria.pdf.

Table 1
Intelligences evaluated, activities used, abilities measured in each intelligence, and the specific items that are rated in each of the abilities.

Intelligence	Activities	Abilities	Items rated
Naturalistic	- Discovery area - Sink and float	Identification of relationships (n1)	1. Notices similarities and dissimilarities between objects. 2. Classifies objects according to various criteria:
		Close observation (n2)	1. Engages in close observation of materials using one or more of the senses, and describes the following characteristics of the objects. 2. Notices changes in the following characteristics of an object over time. 3. Shows interest in recording observations.
		Hypothesis formation (n3)	1. Makes predictions based on observations. 2. Asks "what if"-type questions about natural objects or events. 3. Offers explanations for why things are the way they are.
		Experimentation (n4)	1. Follows up on hypotheses by generating ideas for experiments or setting up simple experimental situations. 2. Explores objects or relationships by manipulating pertinent variables or combining material in ways that are novel.
		Interest in nature activities (n5)	1. Shows interest in natural phenomena or related material over extended period of time 2. Asks questions regularly about things he/she has observed 3. Likes to report on his/her own or other's experiences with the natural environment.
		Knowledge of the natural world (n6)	1. Demonstrates an unusual amount of knowledge about a particular natural object or phenomenon 2. Spontaneously offers relevant information about various natural phenomena
Bodily-kinesthetic [a]	Creative movement	Sensitivity to rhythm (bk1)	1. Attempts to move with the rhythm set by the rater on a drum. 2. Can set a rhythm of her own through movement and regulate it to achieve the desired effects.
		Expressiveness (bk2)	1. Is comfortable using gestures and body postures to express his/herself. 2. Responds vividly to different verbal images. 3. Varies response to music selections, interpreting the quality of music in his/her movements.
		Body control (bk3)	1. Can sequence, and execute movements efficiently. 2. Accurately executes movement ideas proposed by adults or other children. 3. Identifies and uses different body parts and understands their functions. 4. Replicates her own movements and those of others.
		Generation of movement ideas (bk4)	1. Responds immediately to ideas and images with original interpretation.
Spatial [b]	Art portfolio	Level of representation (s1)	1. Basic shapes. 2. Colour. 3. Spatial integration.
		Degree of exploration (s2)	1. Colour. 2. Variants. 3. Dynamics.
		Level of artistry (s3)	1. Expressivity. 2. Repleteness. 3. Aesthetic sensibility.
Musical	Singing	Rhythm (m1)	1. Number of units. 2. Grouping. 3. Pulse. 4. Clarity.
		Pitch (m2)	1. Cadence. 2. Distinction between phrases. 3. Interval. 4. Proper pitch.
		General musical ability (m3)	1. Exceptional production. 2. Expressivenesss.
Logical–Mathematical	Dinosaur game	Numerical reasoning (lm1)	1. Counts correctly up to 6 moving along the corresponding squares.
		Spatial reasoning (lm2)	1. Accuracy of directionality.

(continued on next page)

Table 1 (*continued*)

Intelligence	Activities	Abilities	Items rated
Logical-Mathematical		Logical reasoning (lm3)	1. Recognizes the symbols + and − on the dice, and knows that the former means go forward and the latter go back. 2. Uses the symbols correctly.
Linguistic	- Storyboard - Reporter	Primary language functions (l1)	1. Storytelling. 2. Interacting with adult. 3. Investigating. 4. Labeling or categorizing.
		Narration (l2)	1. Nature of narrative structure. 2. Thematic coherence. 3. Use of narrative voice. 4. Use of dialogue. 5. Use of temporal markers. 6. Expressiveness.
		Information (l3)	1. Level of vocabulary. 2. Sentence structure.

[a] Bodily-kinesthetic intelligence is similar to psychomotor abilities (*Gp)*,kinesthetic abilities (*Gk*), (Carroll, 1993; McGrew, 2009), and psychomotor skills (Fleishman, 1972).

[b] The spatial intelligence measures used in Project Spectrum and in this study are not quite the same as spatial ability as usually tested and defined: "The ability to generate, retain, retrieve, and transform well-structured visual images" (Lohman, 1994, p.1000). This broad ability represents a collection of different abilities, each of which emphasizes a different process involved in the generation, storage, retrieval and transformation of visual images (McGrew, 2009). However, in our study, like in Project Spectrum, the measures used are mainly focussed on evaluating a sense of the "whole" of a subject, a "gestalt" organization (Gardner, 1993).

meeting was held with teachers and parents to explain the study to them and obtain written consent for the children to take part in the research.

The raters were postgraduate students who, during a seminar on multiple intelligences, were given training in how to administer the assessment, and guidelines were provided with respect to the typical behaviors for each item (Plucker et al., 1996; Udall & Passe, 1993). The raters were randomly assigned to the different classrooms so that each rater worked in three different classrooms (either as elicitor and rater or as observer and rater).

Scoring rubrics were developed by project staff before administering the assessment and evaluating the pupils' performance (Plucker et al., 1996).

The activities and assessments were conducted in the different schools over the course of the school year. The indications given in the calendar for the Spectrum class were followed. This calendar provides a schedule for carrying out the different activities throughout the school year (see Krechevsky, 1998) and each pupil was evaluated on the days established for each activity. At the scheduled time, each pupil performed each activity individually so all the children performed all the activities. While the children were performing the activities, the abilities involved in each activity were evaluated using the above scale. For example, rating the information ability (linguistic intelligence) as never, means that the pupil did not demonstrate, while carrying out the activities in question, that he had an adequate level of vocabulary or of sentence structure.

The reliability of the ratings of the abilities measured was established using inter-rater procedures. To test inter-rater reliability, two raters who were experienced in the use of the assessment tool rated the children simultaneously during the ability assessment. Rater 1 assessed the items making up each ability through direct elicitation and observation; rater 2 observed simultaneously and scored the performance independently. Each rater assigned a score to the children's performance (on a scale of 1 to 4). The same rater did not always act as an elicitor or as an observer but changed roles for the different activities so that the measurements were not affected by elicitation style or whether the elicitor liked the particular child in question.

Inter-rater reliability was calculated based on the scores obtained for each item using the Pearson product moment correlation coefficient, r (Stine, 1989; Zegers & ten Berger, 1985). The results showed that r values ranged from 0.78 to 0.97 ($p<.001$), demonstrating good inter-rater reliability for the 60 items measured.

The average score for each of the items assessed was then calculated by taking the mean of the scores given by the two raters. The final score for each ability was obtained by adding up the average scores for the items corresponding to this ability and calculating the mean.

In this study, inter-rater reliability rather than inter-tester reliability (Liao & Pan, 2005) was examined. One rater conducted the procedure, eliciting the child's response and then giving a score. At the same time, the second rater also observed the response and gave an independent score. Hence, most of the variance in the scores probably resulted from the raters' judgment. If inter-tester reliability was to be assessed, both raters would need to administer the test and award scores for the same child. The error variances could then be a consequence of the testing method used, environment, time, and/or raters' judgment.

1.4. Design and data analysis

The data underwent confirmatory factor analysis in order to contrast the proposed theoretical models. All of the confirmatory factor analyses were carried out with AMOS 7 (Arbuckle, 2006). Hierarchical confirmatory factor analysis was done, and nested models were compared to evaluate the structure of the data (Bentler & Dudgeon, 1996; Rindskopf & Rose, 1988).

The use of nested model comparisons in conjunction with hierarchical confirmatory factor analysis extends the confirmatory technique because it allows specific features of the model to be tested. This testing is done by constraining the

Table 2
Intercorrelations between variables.

	n1	n2	n3	n4	n5	n6	bk1	bk2	bk3	bk4	s1	s2	s3	m1	m2	m3	lm1	lm2	lm3	l1	l2	l3
(n = 393)																						
n1	1																					
n2	.401**	1																				
n3	.397**	.555**	1																			
n4	.424**	.441**	.620**	1																		
n5	.593**	.485**	.573**	.690**	1																	
n6	.594**	.513**	.632**	.712**	.797**	1																
bk1	.232**	.191**	.180**	.155**	.241**	.182**	1															
bk2	.143**	.214**	.120*	.119*	.155**	.110*	.496**	1														
bk3	−.042	.049	−.011	−.090	−.095	−.099	.306**	.356**	1													
bk4	.272**	.256**	.224**	.282**	.323**	.342**	.285**	.310**	.124*	1												
s1	.251**	.117*	.161**	.243**	.333**	.236**	.279**	.254**	.052	.351**	1											
s2	.271**	.133**	.103*	.201**	.344**	.249**	.239**	.248**	.036	.346**	.743**	1										
s3	.252**	.102*	.178**	.260**	.346**	.238**	.271**	.256**	−.003	.298**	.713**	.714**	1									
m1	.053	.051	.072	.102*	.025	.049	.049	.106*	.094	.063	.101*	.029	.075	1								
m2	.109*	.081	.158**	.130**	.159**	.159**	.253**	.235**	.190**	.153**	.147**	.113*	.134**	.569**	1							
m3	.280**	.006	.057	.235**	.402**	.283**	.316**	.252**	.017	.307**	.555**	.515**	.500**	.115*	.298**	1						
lm1	.103*	.200**	.131**	.116*	.143**	.123*	.122*	.180**	.074	.206**	.230**	.189**	.149**	.127*	.162**	.075	1					
lm2	.185**	.075	.107*	.212**	.302**	.209**	.184**	.236**	.152**	.286**	.499**	.420**	.384**	.135**	.227**	.466**	.465**	1				
lm3	.186**	.082	.127*	.215**	.315**	.218**	.191**	.222**	.103*	.273**	.481**	.437**	.408**	.141**	.178**	.477**	.434**	.674**	1			
l1	.161**	.055	.098	.138**	.166**	.153**	.156**	.192**	.013	.132**	.170**	.203**	.154**	.080	.123*	.180**	.075	.130**	.119*	1		
l2	.217**	.112*	.219**	.239**	.189**	.201**	.168**	.142**	.053	.149**	.094	.145**	.159**	.121*	.162**	.131**	.075	.092	.097	.676**	1	
l3	.173**	.147**	.172**	.243**	.248**	.198**	.165**	.237**	.024	.212**	.252**	.257**	.256**	.068	.094	.245**	.206**	.293**	.265**	.238**	.210**	1
Mean	2.14	2.11	2.44	2.01	2.15	1.99	2.49	2.54	2.80	2.20	2.16	2.10	2.09	2.82	2.36	.99	3.05	2.87	3.11	1.97	1.87	1.96
SD	.785	.802	.768	.862	.915	.853	.796	.752	.759	.788	.745	.715	.747	.985	1.05	1.16	1.16	1.18	1.01	.603	.615	.647

Note. n1 = Identification of relationships; n2 = Close observation; n3 = Hypothesis formation; n4 = Experimentation; n5 = Interest in nature activities; n6 = Knowledge of the natural world; bk1 = Sensitivity to rhythm; bk2 = Expressiveness; bk3 = Body control; bk4 = Generation of movement ideas; s1 = Level of representation; s2 = Degree of exploration; s3 = Level of artistry; m1 = Rhythm; m2 = Pitch; m3 = General musical ability; lm1 = Numerical reasoning; lm2 = Spatial reasoning; lm3 = Logical reasoning; l1 = Primary Language Functions; l2 = Narration ability; l3 = Information ability.

* $p < .05$, two-tailed.

** $p < .01$, two-tailed.

Table 3
Item-total statistic and reliability coefficients for the six intelligence scales.

Intelligence/ Items	α	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's alpha if item deleted
Naturalistic	.88				
n1		10.70	12.077	.587	.884
n2		10.73	12.038	.577	.885
n3		10.40	11.689	.689	.869
n4		10.83	11.014	.723	.863
n5		10.69	10.356	.798	.850
n6		10.85	10.553	.833	.844
Bodily-kinesthetic	.64				
bk1		7.54	2.683	.507	.515
bk2		7.49	2.705	.553	.485
bk3		7.24	3.137	.344	.630
bk4		7.83	3.143	.313	.653
Linguistic	.63				
l1		3.83	.965	.580	.347
l2		3.92	.969	.552	.384
l3		3.84	1.244	.312	.807
Musical	.72				
m1		3.35	3.190	.410	.457
m2		3.81	2.586	.564	.203
m3		5.18	3.257	.236	.725
Logical-mathematical	.76				
lm1		5.98	4.030	.492	.800
lm2		6.17	3.399	.663	.601
lm3		5.92	4.027	.649	.635
Spatial	.89				
s1		4.19	1.832	.786	.833
s2		4.25	1.906	.787	.832
s3		4.25	1.857	.764	.852

Note. α = Cronbach's alpha.

factorial solution in ways that systematically contrast competing higher-order models. In our study, a six first-order uncorrelated factors model, a six first-order factors and one second-order factor model, and a six first-order correlated factors model are placed in a hierarchy of factor analysis models, which enables the fit of these models to be compared (Rindskopf & Rose, 1988).

Several indices are available to evaluate the adequacy of the models. Bentler (1990) proposed more than thirty. Determination of model fit is not so straightforward; in fact, only some indices, such as chi-square (χ^2) and the RMSEA have an associated statistical test. A number of these fit indices were designed to take into account the size of the sample (Wheaton, 1988). The ratio χ^2/df indicates the size of the comparison statistic χ^2 between the null model and the corresponding model with respect to the number of degrees of freedom. The NNFI (Non Normalized Fit Index) takes into account the degrees of freedom. The PNFI (Parsimony Normalized Fit Index), which is very sensitive to model size, takes parsimony into account, as does Akaike's AIC measure.

Although, the above indices (CFI, NNFI, GFI, and PNFI) were considered conventionally acceptable when values were above .90 (Loehlin, 1998), recent work by Hu and Bentler (1999) suggested a cutoff point of .95 or above for the fit indices.

When the general analysis strategy consists in comparing nested hierarchical models, the improvement in fit is evaluated by the change in χ^2 with respect to the added degrees of freedom. In addition, when models are not nested, the fit of a given model is evaluated by measures based on the population error of approximation (e.g., RMSEA — root mean square error of approximation) and the information measures of fit (e.g., AIC — Akaike's information criterion). Hu and Bentler (1999) suggested RMSEA </= .06 as the cutoff for a good model fit. The test of close fit – PCLOSE – tests the null hypothesis that RMSEA is no greater than .05.

An evaluation of a set of fit indices provides a good idea of the comparative quality of the estimated models. Therefore, here we present a variety of fit indices for the estimated models.

2. Results

2.1. Item analysis and reliability

Classic item analysis was carried out. The aim of this analysis was to evaluate how the measures of each of the abilities are assigned to each intelligence scale, using the correlation between each ability measure and each intelligence scale (Nunnally, 1978). This correlation is used to determine whether an item – ability – belongs to the intelligence scale it has been assigned to, to a different one, or whether it should be eliminated in subsequent analyses.

As we can see in Table 2, almost all items – abilities – belonging to the same scale exhibited high correlations with the items of that particular scale, as compared to items belonging to other intelligence scales. However, there are some exceptions; such as ability m3 (which had a high correlation with abilities s1, s2, s3, lm2, and lm3); bk4 (which showed high correlations with abilities s1 and s2); and l3 (which correlated with several other abilities lm1, lm2, s1, s2,).

Furthermore, the corrected ability–intelligence scale correlations (shown in Table 3) were high, ranging from .31 to .83,

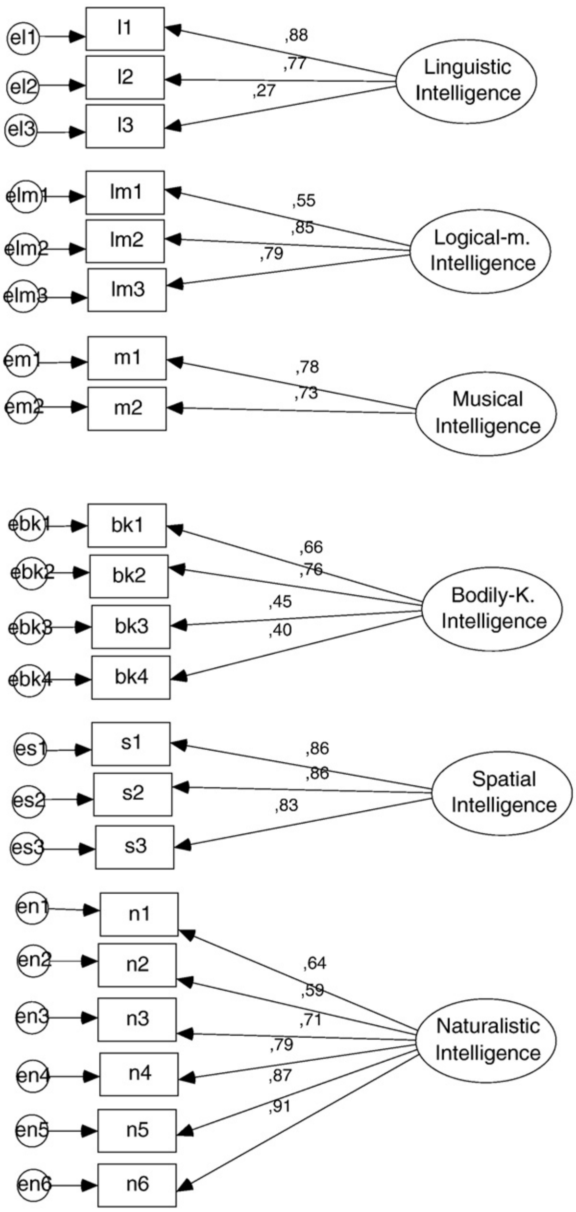

Fig. 1. Model with six first-order uncorrelated factors.

with the exception of ability m3 whose correlation with the musical intelligence scale was .23.

Internal consistency coefficients (Cronbach's α) for each set of abilities belonging to each intelligence scale were calculated. Once alpha had been obtained for the whole intelligence scale, it was recalculated after eliminating one ability in order to verify whether the internal consistency of the scale improved. This effect was particularly marked in the musical scale, where α improved from .58 to .72 when ability m3 was eliminated.

As we can see, the ability that performed worst in all the indices was m3; it displayed high correlations with abilities from other intelligence scales, a lower ability/total scale corrected correlation, and as Table 2 shows, a lower mean score than the other abilities.

This was due to the fact that there were many missing values in the item "exceptional productions" belonging to general musical ability. These missing values were given a score of 0 when, in view of the child's performance, the rater had not scored this item. Although this may not be the best approach to scoring measured ability m3, this score was used in subsequent analyses so as to keep this ability in the study. Nevertheless, in the light of the results, m3 was excluded from confirmatory factor analysis

Table 3 also shows the internal consistency reliability indices for each intelligence scale. The coefficient values are acceptable. In this respect, Nunnally (1978) suggested that values greater than or equal to .70 are preferable, but values around .60 (Hair et al., 1995) have been viewed as acceptable. Moreover, the reliability of scales is related to the number of items (Peterson, 1994; Richardson & Kuder, 1939).

2.2. Comparison of models: confirmatory factor analysis

As shown in Table 2, most of the correlations were positive and statistically significant. Most of the values in the correlation matrix are moderate, which rules out the possibility that the correlation matrix is an identity matrix.

The next step in the data analysis involved establishing different structural models, which corresponded to the four theoretical models mentioned above. All of the models were analyzed following the method of maximum likelihood, under the assumption of multivariate normal distributions, since the values of the variables' skewness and kurtosis ranged between ± 1, with the exception of variables lm3, logical reasoning, (kurtosis = 6.23), and lm2, spatial reasoning (kurtosis = 2.83). However, the method of maximum likelihood, used in AMOS 7, is robust for departures from normality, especially if the sample is large and the skewness and kurtosis values are not extreme, i.e., skewness values > I2I and kurtosis > I7I (Browne, 1984; West, Finch, & Curran, 1995). Furthermore, the bootstrap approach implemented in AMOS 7 was used for model comparison.

Model 1, with six uncorrelated factors, shown in Fig. 1 was based on the initial theoretical framework of multiple intelligences. This model contains six first-order uncorrelated factors. During the parameter estimation process, the model was empirically under-identified, which resulted in one negative error variance estimate. The empirical under-identification may be due to various causes and has various solutions (Rindskopf & Rose, 1988), which should however be technically and theoretically supported (Loehlin, 1998).

In this paper, a very common solution was adopted: the common factor loadings of m1 and m2 on the musical factor were constrained to be equal, which, moreover, in this case was technically supported (critical ratio for difference between factor loadings $CR = -1.938 < 1.95$, $p = .05$; i.e., we cannot reject, at the .05 level, the hypothesis that the two factor loadings are equal in the population) and theoretically supported because m1 and m2 are thought to measure the same musical factor. This modification was included in the other three models.

Fit indices corresponding to Model 1 are shown in Table 4. Taken together, these fit indices did not adequately account for the variance in the matrix. The substantive fit indices (e.g., NNFI), comparative fit indices (e.g., CFI), and the RMSEA did

Table 4
Goodness of fit summary.

Model	χ^2	df	χ^2/df	GFI	CFI	RMSEA	PCLOSE	NNFI	PNFI	AIC
1	737.705	190	3.883	.821	.841	.086	.000	.824	.722	819.705
2	449.962	184	2.445	.899	.923	.061	.007	.912	.768	543.962
3	445.093	183	2.432	.900	.924	.060	.008	.912	.765	541.093
4	425.305	175	2.430	.905	.927	.060	.010	.913	.736	537.305

not reach the values indicated for good fit, in fact they fell well below these values.

Model 2, as shown in Fig. 2, contains six first-order factors and one second-order general factor that underlies all the first-order factors. This model corresponds to the traditional theoretical concept of general and hierarchical factorial intelligence, which postulates the existence of several first-order factors and a prominent second-order factor, g, at the top. This model showed a better fit than did the model with six first-order uncorrelated factors, Model 1. All fit indices were nearer to expected standard values than in Model 1. In particular, indices χ^2/df, and CFI were closer to the cutoff criteria.

Model 3 is similar to the approach used by Visser et al. (2006a). This is a model, with two correlated second-order general factors, in which the cognitive intelligences (linguistic, spatial, logical-mathematical, and naturalistic) load on a general "cognitive" factor *g*, and the non-cognitive intelligences (bodily-kinesthetic, and musical) load on a general "non-cognitive" factor.[2] A path diagram depicting this structure is provided in Fig. 3.

In this model, the cognitive and non-cognitive factors correlated .73 (SE = .10). This model showed a better fit than did the preceding models, the model with six first-order uncorrelated factors, Model 1, and the model with one second-order general factor that underlies all the six first-order factors, Model 2. Fit indices for Model 3, the two correlated second-order general factors model, were closer to commonly accepted values. Nevertheless, a fourth model was also tested.

Model 4, a model with six first-order correlated factors, represents a modification of Model 1 in that it allowed the latent intelligence variables to correlate. Fig. 4 represents the model.

An examination of the fit indices presented in Table 4 suggests that the model with six first-order correlated factors, Model 4, showed a reasonable fit to empirical data. The ratio χ^2/df was very acceptable (χ^2/df = 2.43); the values of the Comparative Fit Index, CFI = .927, and the RMSEA = .060, were within the accepted values of goodness of fit. Overall, the fit indices of Model 4 were slightly better than those of Model 3, with the exception of the PNFI-Parsimony Normalized Fit Index.

Because these four models are nested, a chi-square statistic could be computed to compare each of the solutions directly (i.e. change in χ^2 with respect to the added degree of freedom). The results of this comparison are shown in Table 5. In general, all the models fitted significantly better than the preceding models and the model with six first-order correlated factors (Model 4) fitted statistically better than the model with two correlated general factors (Model 3), (χ^2 difference = 19.78, df = 8, $p < .02$).

Furthermore, given that some variables had a minor to moderate departure from normality, bootstrapping was used for model selection, with the original sample serving as the population for the purpose of bootstrap sampling. The bootstrapping procedure for model comparison – with 1000 bootstrap samples, of which 0 bootstrap samples were unused – showed that the lowest mean discrepancy occurred for Model 4, the model with six first-order correlated factors.

Of the four models compared, the two with the best fit to empirical data were Model 3 and Model 4. The differences in fit between these two models were statistically significant, but only at a low level of significance. This requires a close inspection of the difference between Model 3 – the "cognitive" and "non-cognitive" second-order correlated general factors model – and Model 4, the model with six first-order correlated factors.

Model 4, the model with six first-order correlated factors, might be expected to have a better fit simply because it has fewer restrictions. The problem with simple fit indices is that, within any series of models, the fit always improves as one solves for more complex or less restricted models (Loehlin, 1998). Therefore, a number of indices penalize complexity, such as the Parsimony Normalized Fit Index (PNFI), proposed by James, Mulaik, and Brett (1982). This index, although very sensitive to model complexity, was better for Model 3, and favored this model with two correlated second-order general factors, the "cognitive" and "non-cognitive" factors, over Model 4.

On the other hand, the parameters estimated for Model 4 revealed low correlations between the first-order factors. The values of the correlations ranged from .61 (logical/mathematical-spatial) to .16 (musical-spatial), (see Fig. 4) with a median value of .30.

Nevertheless, it should be pointed out that no model exhibited a totally satisfactory fit to the empirical data; that is, the value of the chi-square statistic and the P-value for the test of close fit PCLOSE yielded a probability $p < .05$. The fit of a model may be improved introducing modifications in certain parameters. However, these modifications must be based on a theoretical model of intelligence. Therefore, a theoretically justifiable model that provides a less satisfactory fit than a model adjusted on an *ad hoc* – or data driven – basis should be preferred (Dunn, Everitt, & Pickles, 1993).

2.3. Variance Decomposition: Contribution of g and Intelligences to Ability Measures

As mentioned earlier, confirmatory factor analysis allows us to obtain estimates of the influence of the general factor on

[2] This model was suggested by a reviewer of this paper as a modification of a previous model proposed by the authors which did not allow cognitive and non-cognitive general factors to correlate. However, the model proposed initially gave a poorer fit than did the current Model 3.

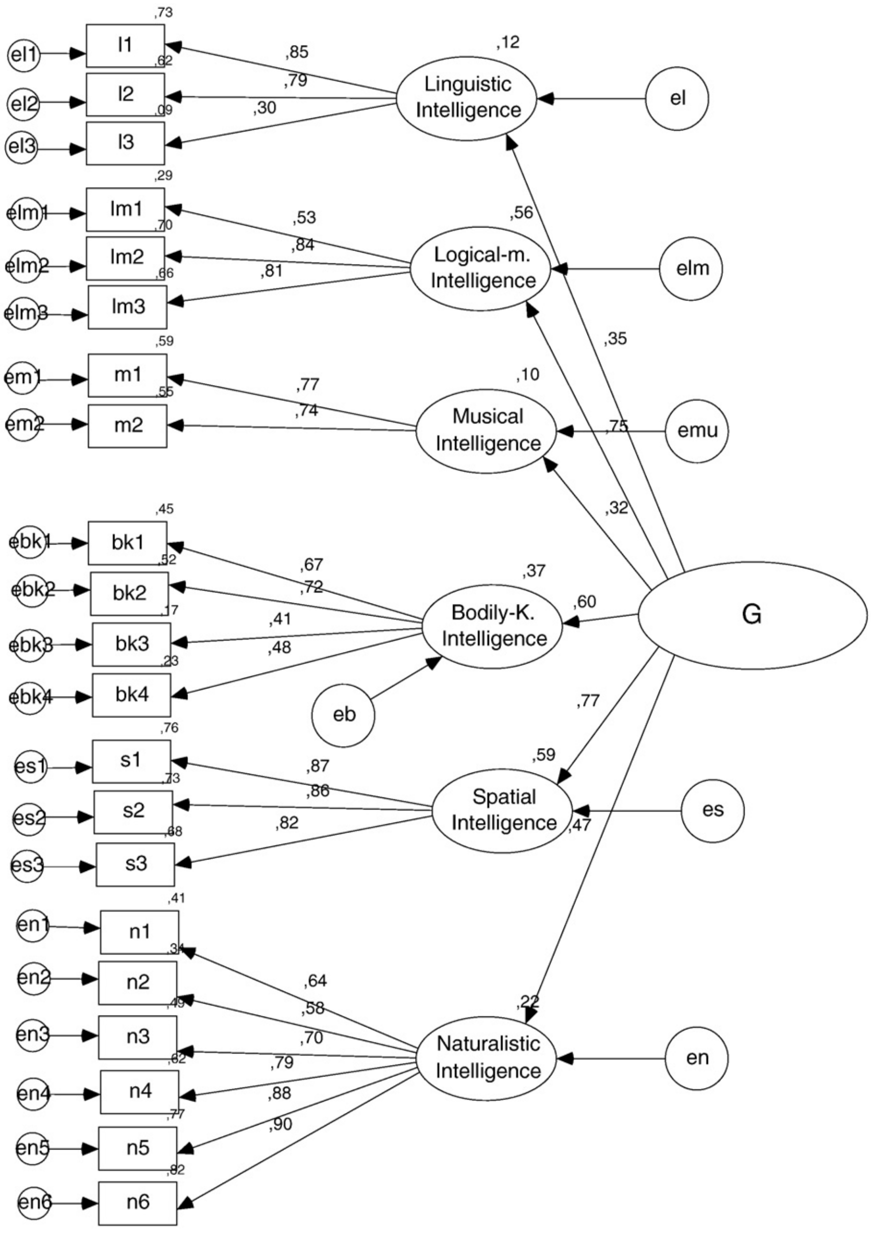

Fig. 2. Model with six first-order factors, and one second-order general factor.

each of the abilities by calculating the indirect effects of the second-order general factor on measured abilities. The second-order factor model can be used to get separate estimates of the second-order factor- *g*-, first-order factors, and unique/error variances for each measure.

Table 6 shows the variance decomposition in Model 2, with a single second-order general factor.

Column 3 contains the percentage of variance in each measure of ability that is attributable to *g* (% variance explained by *g*), calculated by squaring the values of the indirect effect (column 2) of *g* for individual abilities. Column 4 contains the percentage of variance explained by first-order factors (intelligences). Meanwhile, column 5 shows the percentage of variance explained by *g* and first-order factors. In column 6 the percentage of unexplained variance in each measured ability is shown.

As we can see the percentage of variance due to *g* is different for different abilities. Some measured abilities were more highly *g*-loaded than others, as could be expected. However, contrary to what might be expected, the abilities of linguistic and naturalistic intelligence were low *g*-loaded.

Comparing the percentage of variance due to *g* with variance explained by first-order factors – intelligences – we can see that of the 21 measured abilities, the variance in 15 abilities was due to specific first-order factors more than to *g* variance. On average, *g* explained less variance (17.9%) in the abilities than did first-order factors or more specific intelligences (35.4%). Of the total percentage of explained variance

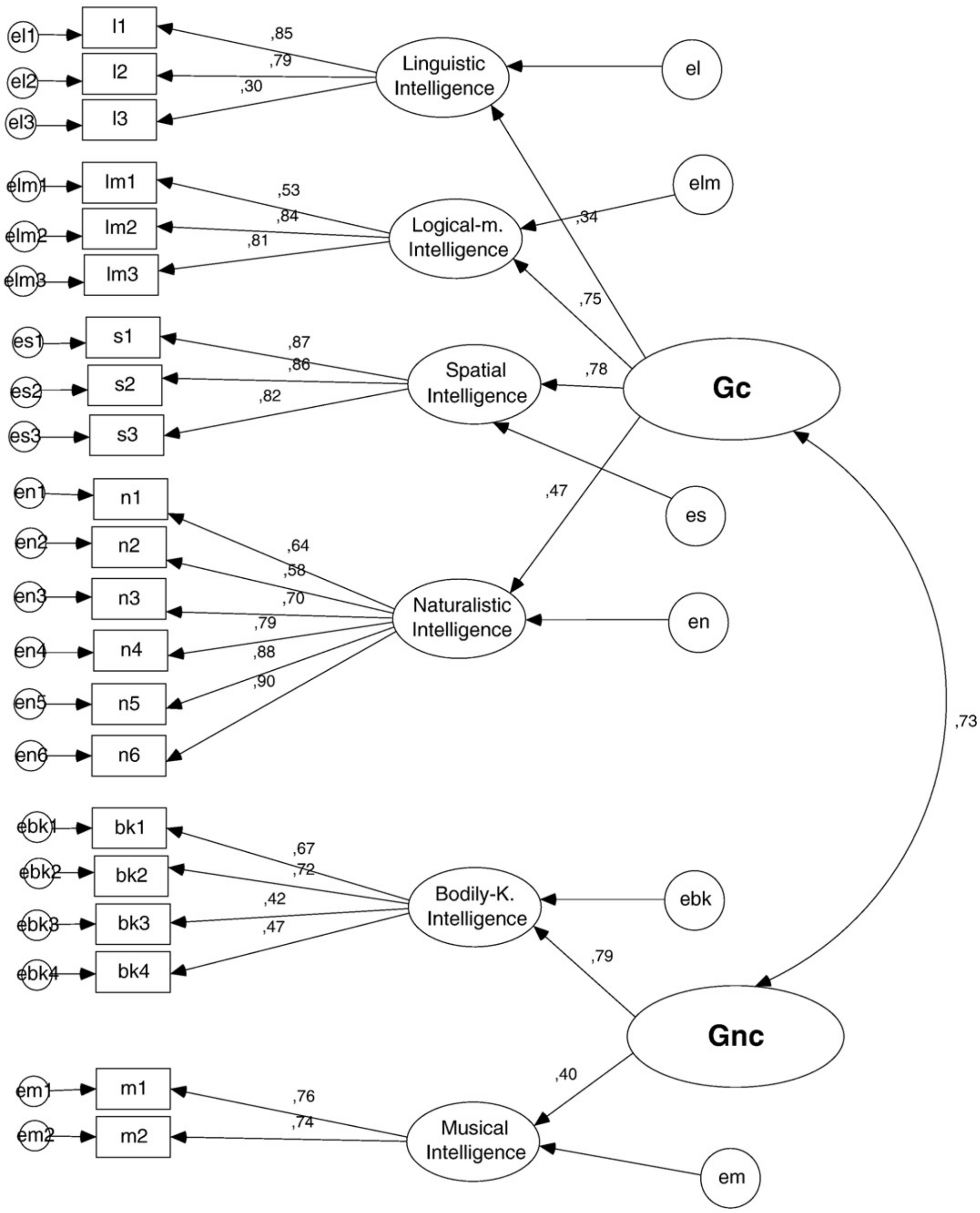

Fig. 3. Model with six first-order factors, and two correlated second-order general factors.

in the abilities, 66% corresponded to first-order factors and 33% to *g*.

With regard to total explained variance, several measures such as n6, n5, s1, s2, l1, lm2, etc. had quite acceptable values of .70 or higher. However, it is important to point out that various measures had very low levels of explained variance, as shown in Table 6. The abilities with lower values of explained variance were l3, bk3, bk4, lm1, and n2. Since these abilities had poor levels of explained variance, they could not have much *g* variance or first-order factor specific variance.

Nevertheless, the average proportion of residual (unexplained) variance in the abilities, a combination of unique and error variance, was 46.6%, which was less than the average proportion of variance in the abilities explained by *g* and first-order factors, 53.4%.

3. Discussion

In relation to the goodness of fit, an increase was observed in the model with a single second-order general factor (Model 2) as compared with the model of six first-order uncorrelated factors (Model 1). Clearly, Gardner's intelligences are not quite independent of each other. In addition, Model 3, with two correlated second-order general factors,

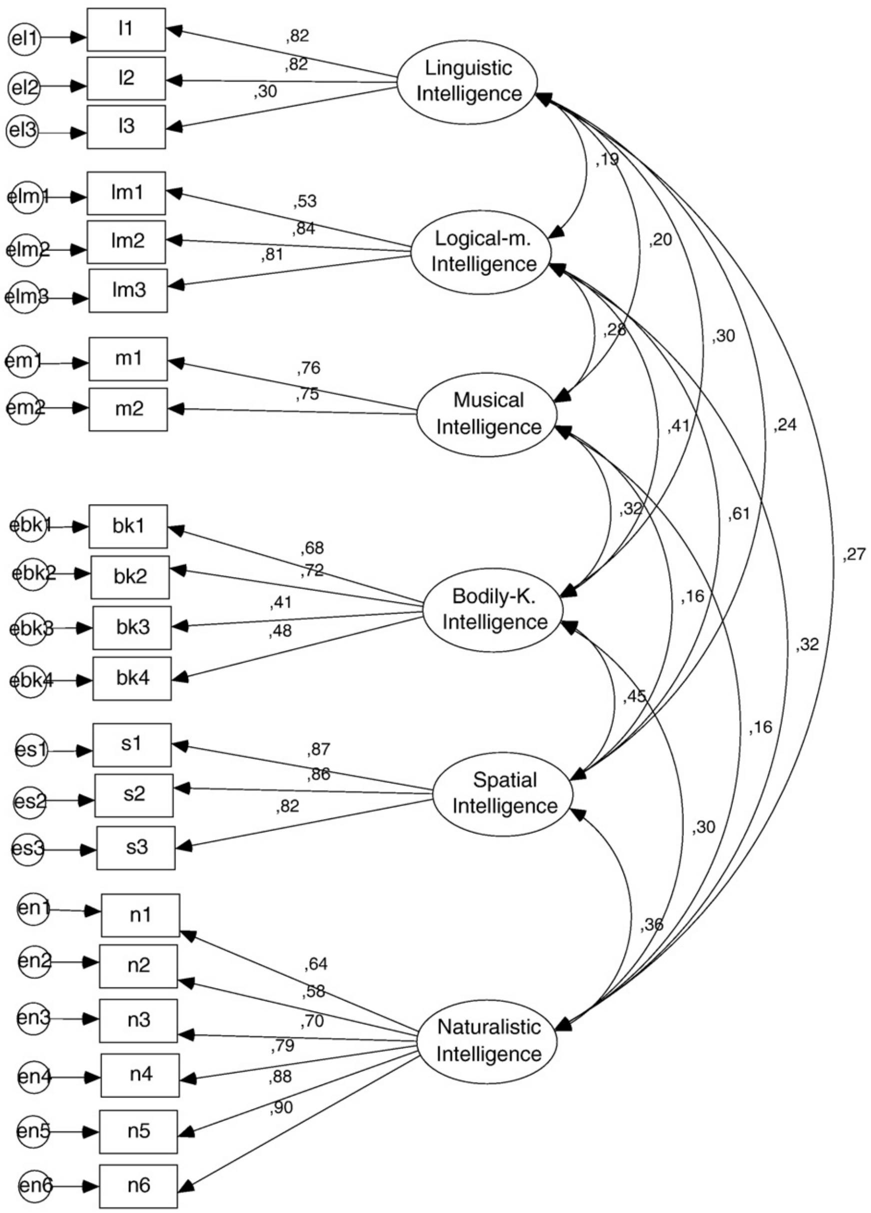

Fig. 4. Model with six first-order correlated factors.

the "cognitive" and "non-cognitive" factors, fitted significantly better than did Model 2, with a single general second-order factor, the *g* factor (Carroll, 1993). This was congruent with the results obtained by Visser et al. (2006a). Model 3 is the model that intelligence researchers may find attractive when they consider Gardner's theory. It might be a strong competitor to Gardner's model. In fact, Carroll (1993) pointed out that Gardner's intelligences bear a striking similarity to the second-stratum factors of Carroll's hierarchy. However, psychomotor ability is not recognized as an aspect of cognitive ability and, thus, bodily-kinesthetic intelligence appears to have no counterpart in second-stratum factors, whereas musical intelligence has a sensory as well as a cognitive component. In Model 3, cognitive and non-cognitive general factors correlated .73, which differs considerably from 1. This result is congruent with the distinction between two categories of intelligences in Gardner's model according to the extent to which cognitive and non-cognitive – or partly non-cognitive – abilities were included in each general factor. Although Visser's model might be expected to fit better, the two correlated general factors solution model did not fit as well as Model 4, the six first-order correlated factor model, as regards most of the indices, with the exception of PNFI, an index that penalizes complexity.

Model 3 and Model 4 were the models that showed a better fit to empirical data. The differences between Model 3 – a model with two correlated general factors, "cognitive" and "non-cognitive" factors – and Model 4 – a model with six first-order correlated factors, corresponding to a weaker but more recent version of the theory of multiple intelligences (Gardner, 2006) –

Table 5
Chi-square test for the change in fit between nested models.

Nested models	χ^2 difference	df difference	p
Model 1. Six first-order uncorrelated factors	–	–	–
Model 2. Six first-order factors, and one second-order general factor	287.743	6	.001
Model 3. Six first-order factors, and two correlated second-order general factors	4.869	1	.04
Model 4. Six first-order correlated factors	19.788	8	.02

were small. Model 4 fitted slightly better than Model 3; a statistically significant increase in fit was observed in the model with six correlated factors. Furthermore, the first-order factor correlations in Model 4 were moderate or low, indicating that intelligences are relatively independent. The lack of substantial correlations among the latent first-order factors implies that *g* variance in the measured abilities (i.e., the shared relationships among the independent latent factors) was lower than the components of variance in the latent factors that were independent of each other.

Although the analyses do not rule out the presence of *g* variance in these abilities, the percentage of variance due to first-order factors was higher. Furthermore, the loading of the first-order factors on *g* in Model 2 was moderate (from .77 for spatial to .32 for musical intelligence). With these loadings, the general factor only accounted for 59% to 10% of the variance. This questions the existence of a general factor that underlies these abilities, although these loadings are also representative of many factor analyses with second-order factors such as Carroll's study. In fact, although Model 4 fitted better than Model 2, when viewed from the perspective of parsimony, preference is given to the simpler solution — the one second-order *g* factor model, Model 2.

From the theoretical perspective, the model with six first-order correlated factors, Model 4, can be assimilated into the theory of multiple intelligences if we take into account two assumptions of this theory: a) the assumption that all individuals possess all of the intelligences, but differ in relative strengths and weaknesses; and b) intelligence never exists in isolation from other intelligences (Gardner, 2003; Krechevsky & Seidel, 1998). In other words: “Although Gardner has diluted MI theory somewhat by incorporating the existence of *g* and suggesting that the intelligences might not be entirely independent, the theory would still seem to predict that tests of the eight intelligences should be relatively independent of each other” (Visser et al., 2006a, p. 492).

Model 4 could also be congruent with the Cattell–Horn–Carroll (CHC) theory. The updated Horn–Carroll theory (Horn, 2007; McGrew, 2005) has about 9 higher-order factors, including crystallized intelligence (similar to Gardner's linguistic intelligence), fluid intelligence (similar to Gardner's logical intelligence), general spatial intelligence (like Gardner's spatial intelligence), general auditory intelligence (could be similar to Gardner's musical intelligence), etc. These broad factors, although positively correlated, are operationally independent and have predictive independence, as well as independence in virtue of having distinct construct validities. Horn (2007) did not postulate that it was necessary to hypothesize a general *g* factor above the nine higher-order factors, so Horn's theory is in line with this aspect of Gardner's theory.

An additional aspect studied was the influence of the general factor, *g*, and the first-order factors – intelligences in Gardner's view – on measured abilities. The results obtained in our study did not altogether favor a single theory. On the one hand, the effect of *g* was greater in the abilities that are assumed to be more *g*-loaded, such as those belonging to spatial and logical/mathematical intelligence, and less in the abilities assumed to be less *g*-loaded, such as the abilities of musical and bodily-kinesthetic intelligence. This is in agreement with the hypothesis of Visser et al. (2006a) and with what was to be expected in our study also. Visser's findings also showed that, in some cases, “the residual correlations were large enough to suggest a considerable influence of non-*g* sources of variance on the relations between the tests” (Visser et al., 2006a, p. 499). This scarce influence of *g* suggests that the individual tasks do tap into abilities other than *g*. These results are similar to Gridley's (2002) findings and suggest that these tasks measure something more than general intelligence.

Visser et al. (2006b), when discussing the relations between multiple intelligences and *g*, indicated that each of the domains proposed by Gardner appears to involve a blend of *g*, cognitive abilities other than *g* (group factors), and, in some cases, non-cognitive abilities. This pattern of results could be consistent with the *g* + specific-achievement CHC framework, using structural equation modeling techniques, in the claim that both broad and narrow or specific cognitive abilities make important contributions to understanding specific academic abilities, over and above the variance accounted for by *g*; e.g., in reading decoding skills (Floyd, Keith, Taub, & McGrew, 2007), and reading comprehension (Benson, 2008).

When all the analyses are viewed together, along with the goodness of fit indices, the nested comparison models, analysis of percentages of variance due to *g* and the first-order factor intelligences, “it would seem that the Spectrum activities are not so separate from general ability as proposed by the original authors, nor so unitary as argued by their critics” (Gridley, 2002, p. 9).

This study is subject to certain limitations: 1) the moderate reliability of the intelligence scales, made up of the whole set of specific abilities that evaluate each of the intelligences considered in this study, which, nevertheless, are within commonly accepted values; 2) the low reliability of some measures of the Project Spectrum abilities, which use performance-based assessment, an “alternative assessment” not without controversy (Plucker et al., 1996) because of the relative subjectivity of this procedure. Therefore, if some of these abilities had poor levels of reliability they could not have much common *g* variance or specific variance; furthermore, the measures could be affected by a certain kind of rater bias — the tendency to give a particular child a similar rating across all measures for a given construct[3]; 3) the different *g*-loading in the extent to which subtest constituents – abilities – were correlated with each other and had a

[3] This possibility, which could affect the validity of the measures, was suggested by a reviewer of this paper.

Table 6
Standardized indirect effect estimates of general second-order factors on abilities in Model 2, and variance explained in the measured abilities by *g* and first-order factors (intelligences).

Ability	Indirect effect of *g*	% variance explained by *g*	% variance explained by first-order factors (intelligences)	% variance explained by *g* and first-oder factors	% unexplained variance (unique + error variance)
n5	.414	.171	.601	.772	.228
n4	.371	.138	.480	.618	.382
n3	.332	.110	.384	.494	.506
n2	.275	.076	.265	.341	.659
n1	.302	.091	.320	.411	.589
s2	.658	.433	.298	.731	.269
s3	.633	.401	.276	.677	.323
s1	.671	.450	.311	.761	.239
bk2	.436	.190	.331	.521	.479
lm1	.399	.159	.126	.285	.715
lm3	.606	.367	.290	.657	.343
lm2	.627	.393	.310	.703	.297
m2	.239	.057	.490	.547	.453
m1	.248	.062	.527	.589	.411
bk4	.289	.084	.145	.229	.771
l3	.103	.011	.077	.088	.912
l2	.274	.075	.546	.621	.379
l1	.296	.088	.638	.726	.274
n6	.426	.181	.634	.815	.185
bk3	.247	.061	.106	.167	.833
bk1	.407	.166	.287	.453	.547

Note. n5 = Interest in nature activities; n4 = Experimentation; n3 = Hypothesis formation; n2 = Close observation; n1 = Identification of relationships; v2 = Degree of exploration; s3 = Level of artistry; s1 = Level of representation; bk2 = Expressiveness; lm1 = Numerical reasoning; lm3 = Logical reasoning; lm2 = Spatial reasoning; m2 = Pitch; m1 = Rhythm; bk4 = Generation of movement ideas; l3 = Information ability; l2 = Narration ability; l1 = Primary language functions; n6 = Knowledge of the natural world; bk3 = Body control; bk1 = Sensitivity to rhythm.

different percentage of g variance, or alternatively a different specific variance, suggests that these intelligences do not belong to a totally coherent ability domain. In this sense, CHC theory could be used both to examine the structure and relations of stratum II (broad or group factors, similar to Gardner's intelligences) with stratum I (narrow or specific abilities, similar to sub-abilities in the Project Spectrum) and to guide the development and interpretation of ability tests (Alfonso, Flanagan, & Radwan, 2005); 4) only two abilities were measured in relation to musical intelligence, which is not considered sufficient in either exploratory (Gorsuch, 1983) or confirmatory factor analysis (Rindskopf & Rose, 1988); 5) the usefulness of having included intrapersonal and interpersonal intelligences, although Project Spectrum does not evaluate these intelligences separately; and 6) the same person who elicited the task also evaluated the children. In an attempt to minimize this limitation we used two raters.

Finally, this study focused solely on the construct validity of Spectrum activities (and the abilities and intelligences included in each of them) through confirmatory factor analysis. More experimental, construct validity and predictive studies are needed to establish the true usefulness of the theory of multiple intelligences.

References

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 185—202)., 2ª ed. New York: Guilford Press.

Arbuckle, J. L. (2006). *(Version 7.0) [Computer Program].* Chicago: SPSS.

Benson, N. (2008). Cattell–Horn–Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, *26*(1), 27—41.

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

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: statistical practice, theory, and directions. *Annual Review of Psychology*, *47*, 563—592.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 16—33). New York: Cambridge University Press.

Brody, N. (2006). Geocentric theory: a valid alternative to Gardner's theory of intelligence. In J. A. Schaler (Ed.), *Howard Gardner under Fire: The Rebel Psychologist Faces his Critics* (pp. 73—94). Chicago and La Salle: Ill: Open Court.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62—83.

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

Ceci, S. J. (1990/1996). *On Intelligence... More or Less: A Biological Treatise on Intellectual Development.* Englewood Cliffs, NJ: Prentice Hall.

Dunn, G., Everitt, B., & Pickles, A. (1993). *Modelling Covariances and Latent Variables Using EQS.* London: Chapman and Hall.

Fleishman, E. A. (1972). Structure and measurement of psychomotor abilities. In R. N. Singer (Ed.), *The Psychomotor Domain: Movement Behavior* (pp. 78—106). Philadelphia, PA: Lea and Febiger.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, *22*(2), 200—233.

Gardner, H. (1983). *Frames of Mind.* New York: Basic Book.

Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences*, 2ª Ed. London: Fontana Press.

Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century.* New York: Basic Books.

Gardner, H. (2003). Three meanings of intelligence. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of Intelligence. International Perspectives* (pp. 43—54). Washington: American Psychological Association.

Gardner, H. (2006). On failing to grasp the core of MI theory: a response to Visser et al.. *Intelligence*, *34*, 503—505.

Gardner, H., Feldman, D., & Krechevsky, M. (1998). *Project Spectrum: Frameworks for Early Childhood Education.* New York: Teachers College Press.

Gorsuch, R. L. (1983). *Factor Analysis*, 2nd ed. Hillsdale, NJ: LEA.
Gridley, B. E. (2002). In search of an elegant solution: reanalysis of Plucker, Callahan, & Tomchin, with respect to Pyryt and Plucker. *Gifted Child Quarterly*, *46*(3), 1−11.
Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis with Readings.* Englewood Cliffs, NJ: Prentice Hall.
Horn, J. L. (2007). Understanding human intelligence: where have we come since Spearman? In R. C. MacCallum, & R. Cudeck (Eds.), *Factor Analysis at 100: Historical Developments and Future Directions* (pp. 205−247). Mahwah, NJ: Lawrence Erlbaum Associates.
Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253−270.
Horn, J. L., & Noll, J. (1994). A system for understanding cognitive capabilities: a theory and the evidence on which it is based. In D. K. Detterman (Ed.), *Current Topics in Human Intelligence. Volume IV: Theories of Intelligence* (pp. 151−203). Norwood, NJ: Ablex.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modelling*, *6*(1), 1−55.
James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal Analysis: Assumptions, Models, and Data.* Beverly Hills, CA: Sage Publications.
Jensen, A. R. (1998). *The g Factor: The science of Mental Ability.* Westport, CT: Praeger/Greenword.
Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: it is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393−416.
Johnson, W., Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: consistent results from five test batteries. *Intelligence*, *36*, 81−95.
Krechevsky, M. (1998). *Project Spectrum: Frameworks for Early Childhood Education*, *Volume 3*, New York: Teachers College Press.
Krechevsky, M., & Seidel, S. (1998). Minds at work: applying multiple intelligences in the classroom. In R. J. Sternberg, & W. M. Williams (Eds.), *Intelligence, Instruction, and Assessment* (pp. 17−42). Mahwah, NJ: LEA.
Liao, H. -F., & Pan, Y. -L. (2005). Test-retest and inter-rater reliability for the comprehensive development inventory for infants and toddlers diagnostic and screening tests. *Early Human Development*, *81*, 927−937.
Loehlin, J. C. (1998). *Latent Variable Models*, 3rd ed. Hillsdale, NJ: LEA.
Lohman, D. F. (1994). Spatial ability. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 1000−1007). New York: McMillan.
McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: past, present and future. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 136−177)., 2ª ed. New York: Guilford Press.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1−10.
Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.
Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*(2), 381−391.
Plucker, J. A., Callahan, C. M., & Tomchin, E. M. (1996). Wherefore art thou, multiple intelligences? Alternative assessment for identifying talent in ethnically diverse and low income students. *Gifted Child Quarterly*, *40*, 81−92.
Pyryt, M. C. (2000). Finding g: easy viewing through higher order factor analysis. *Gifted Child Quarterly*, *44*, 190−192.
Reynolds, M. R., & Keith, T. Z. (2006). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*, doi:10.1016/j.intell.2006.08.002.
Richardson, M. W., & Kuder, G. F. (1939). The calculation of test reliability coefficients based upon the method of rationale equivalence. *Journal of Educational Psychology*, *30*, 681−687.
Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioural Research*, *23*, 51−67.
Spearman, C. (1904). “General intelligence” objectively determined and measured. *American Journal of Psychology*, *15*, 201−293.
Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence.* New York: Cambridge University Press.
Sternberg, R. J. (2003). Issues in the theory and measurement of successful intelligence: a reply to Brody. *Intelligence*, *31*, 331−337.
Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test (multiple-choice items) in three international samples: an empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment*, *17*, 1−16.
Stine, W. W. (1989). Inter-observer relational agreement. *Psychological Bulletin*, *106*, 341−347.
Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.
Udall, A. J., & Passe, M. (1993). *Gardner-based/performance-based assessment notebook.* Charlotte, NC: Charlotte-Mecklenburg Schools.
Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006a). Beyond g: putting multiple intelligences theory to the test. *Intelligence*, *34*, 487−502.
Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006b). *g* and the measurement of Multiple Intelligences: a response to Gardner. *Intelligence*, *34*, 507−510.
West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables. In R. H. Hoyle (Ed.), *Structural Equation Modelling: Concepts, Issues and Applications* (pp. 56−75). Thousand Oaks, CA: Sage.
Wheaton, B. (1988). Assessment of fit in over-identified models with latent variables. In J. S. Long (Ed.), *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research* (pp. 193−225). Beverly Hills, CA: Sage.
Zegers, F. E., & ten Berger, J. M. F. (1985). A family of association coefficients for metric scales. *Psychometrika*, *50*, 17−24.