Available online at www.sciencedirect.com
ScienceDirect
Intelligence 34 (2006) 487 – 502
INTELLIGENCE

Beyond *g*: Putting multiple intelligences theory to the test

Beth A. Visser [a,*], Michael C. Ashton [a], Philip A. Vernon [b]

[a] *Department of Psychology, Brock University, St. Catharines, Ontario, Canada L2S 3A1*
[b] *Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2*

Received 13 December 2005; received in revised form 20 February 2006; accepted 20 February 2006
Available online 6 March 2006

Abstract

We investigated Gardner's "Theory of Multiple Intelligences" in a sample of 200 adults. For each of the hypothesized eight "intelligence" domains–Linguistic, Logical/Mathematical, Spatial, Interpersonal, Intrapersonal, Musical, Bodily-Kinesthetic, Naturalistic–we selected two tests based on Gardner's description of its content. Factor analysis revealed a large *g* factor having substantial loadings for tests assessing purely cognitive abilities–Linguistic, Logical/Mathematical, Spatial, Naturalistic, Interpersonal–but lower loadings for tests of other abilities, especially Bodily-Kinesthetic. Within most domains, the two tests showed some (weak) non-*g* associations, thus providing modest support for the coherence of those domains, which resemble the group factors of hierarchical models of intelligence. Results support previous findings that highly diverse tests of purely cognitive abilities share strong loadings on a factor of general intelligence, and that abilities involving sensory, motor, or personality influences are less strongly *g*-loaded.

Keywords: Intelligence; Multiple intelligences; *g*; Mental abilities; Cognitive abilities

1. Introduction

The notion of general intelligence or *g* (e.g., Spearman, 1927) had long been broadly accepted by psychologists when Howard Gardner introduced Multiple Intelligences (MI) theory in his 1983 book, *Frames of Mind*, proposing that there are several independent ability areas. Gardner (1993) described intelligence as a biopsychological potential that could be influenced by experience, culture, and motivational factors. He defined intelligence as the ability to solve problems and to fashion products that are culturally valued. Gardner (1983) initially proposed that there were seven intelligences: Linguistic, Spatial, Logical/Mathematical, Interpersonal, Intrapersonal, Bodily-Kinesthetic, and Musical. He has more recently added Naturalistic intelligence and has suggested that an Existential intelligence might exist, but that a hypothesized Spiritual intelligence does not (Gardner, 1999).

Gardner (1999) stated that his choice of the word "intelligences" was a deliberate one, noting that if he had written a book referring to "faculties" or "gifts," it is unlikely that his theory would have garnered the attention that it has. Gardner has professed to be quite willing to refer to his eight intelligences as talents or abilities, but only if verbal and quantitative abilities are referred to as talents. Gardner has argued that there is no hierarchy of ability, and that Linguistic and Logical/Mathematical abilities are of no greater real-life importance than any of the other "intelligences".

Gardner (1999) explained that he reviewed hundreds of studies before publishing *Frames of Mind*, and that he

* Corresponding author.
E-mail addresses: b.visser@cogeco.ca (B.A. Visser), mashton@brocku.ca (M.C. Ashton), vernon@uwo.ca (P.A. Vernon).

0160-2896/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2006.02.004

assessed all candidate intelligences on the basis of eight criteria: the potential of isolation by brain damage; an evolutionary history and evolutionary plausibility; an identifiable core operation or set of operations; susceptibility to encoding in a symbol system; a distinct developmental history; the existence of savants, prodigies, and other exceptional people; support from experimental psychological tasks; and support from psychometric findings. Gardner (1983) admitted that the criteria were somewhat flexible; some intelligences which might have met all criteria, such as face-recognition, were discarded because they were not highly valued within cultures. In addition, Gardner retained candidate intelligences which seemed promising even if they did not meet every criterion.

MI theory was enthusiastically welcomed by many educators and parents (Daz-Lefebvre, 2004; Mettetal, Jordan, & Harper, 1997). Here, Gardner's message that children have unique and diverse abilities meshed well with educators' intuitive sense that children learn in very different ways. Indeed, Gardner's premise seemed far more egalitarian than did notions of *g*. MI theory seemed to say that a child who performed poorly in, say, math and reading had just as great a chance as anyone else at being successful in music, art, physical education, or even geography: everyone could be smart in some way. At a time when standardized testing in US schools had become highly controversial, Gardner (1983, p. 3) claimed that IQ tests had little utility in predicting success beyond school. This claim, however, is clearly contradicted by Gottfredson's (2002) finding that no meta-analysis has reported exceptions to the generality of *g* in predicting job performance, with brighter employees always performing better on average than less intelligent employees.

Although Gardner (1999) has acknowledged the existence of *g*, he has continued to question its explanatory power. In addition, he has maintained that each of his intelligence domains has unique processing resources, and that there are no horizontal capacities, such as memory or creativity, that cut across all hypothesized intelligences. Instead, he views creativity as an operation performed within a domain, rather than as a general, cross-cutting ability. Gardner has suggested that executive functioning likely emerges from Intrapersonal intelligence rather than constituting an intelligence of its own or a horizontal capacity. Gardner (1999, p. 106) stated that he had no objection to others invoking an executive function, but that for the purposes of modeling mental abilities, "it is useful to see whether one can explain human behavior in the absence of such hierarchical considerations, or whether the hierarchy can emerge naturally, as part of everyday functioning, rather than by invoking a separate executive intelligence".

MI theory proposes that the eight intelligence domains are theoretically independent, but Gardner (1993) has acknowledged that two or more could overlap. He cautioned, however, that correlations among subtests of standardized intelligence tests occur because the tasks all rely on rapid responses to items that are heavily based on logical/mathematical and linguistic abilities. However, Messick (1992) noted that variability in reading ability should not influence intelligence test performance as long as all participants are able to easily understand the task instructions. Gardner has expressed concern about the "verbal lens"–that is, the use of a common verbal format to assess all aspects of intelligence–but Messick claimed that the reasoning component of the Logical/Mathematical domain is a far more ubiquitous element across tests of cognitive ability. Messick noted that the reasoning that a person employs to solve a novel task in an intelligence domain other than Logical/Mathematical appears more similar to a horizontal, cross-cutting ability than to method variance. In a similar vein, Lohman (2001) argued that inductive reasoning, with its component of central working memory, was equivalent to *g*.

The content of Gardner's intelligence domains suggests some similarities to the group factors of hierarchical models of intelligence (e.g., Vernon, 1961), and Carroll (1993) has pointed out that Gardner's intelligences bear a striking similarity to the second-stratum factors of Carroll's hierarchy. For example, Carroll noted that Gardner's Linguistic intelligence corresponded to the factor of crystallized intelligence, Musical intelligence to auditory perception ability, Logical/Mathematical intelligence to fluid intelligence, and Spatial intelligence to visual perception. Interpersonal or social abilities, in Carroll's framework, were represented to some extent in first-stratum factors of knowledge of behavioral content (with separate factors emerging for convergent and divergent tasks assessing those abilities). Carroll stated that only Gardner's Bodily-Kinesthetic and Intrapersonal intelligences appeared to have no counterpart in second-stratum factors. However, psychomotor ability is not typically recognized as an aspect of cognitive ability and, thus, Bodily-Kinesthetic ability would not be represented in hierarchical models. Carroll noted that adequate measures of intrapersonal ability have never been included in factor analytic studies of cognitive structure. Gardner had not introduced his eighth domain, Naturalistic intelligence, at the time of Carroll's writing, but the categorization of objects would seem to be related to logical reasoning.

2. Assessment of multiple intelligences

Gardner (1999) has claimed that Linguistic and Logical/Mathematical intelligences are the domains most highly valued in school and most heavily represented on traditional intelligence tests. The challenge, he has said, is to develop assessment tools that are "intelligence fair"; that is, measures that do not assess intelligence through the lens of verbal, or any other, ability. A difficulty with the measurement of the eight hypothesized intelligence domains is Gardner's (1999) argument that assessments should reflect an individual's success in completing culturally valued tasks, rather than in completing intelligence tests. Given Gardner's contention that the performance of real-life activities necessarily employs the combination of two or more "intelligences", the assessment of the individual intelligences of his model would seem to be an impossible task. For example, Gardner has noted that in his research with preschoolers, he has assessed Logical/Mathematical, Spatial, and Bodily-Kinesthetic abilities by having children take apart and reassemble pencil sharpeners and doorknobs. However, Gardner has not explained how each of these three abilities could be assessed independently by such a test.

Gardner (1993) stated that he and colleagues were engaged in efforts to create operational definitions and diagnostic procedures for each intelligence area, although he admitted that it could be difficult to define and assess Intrapersonal and Interpersonal intelligences. In 2006, there still does not appear to be any standardized testing instrument for the multiple intelligences, and Gardner has provided little guidance as to how his intelligences might be tested. A few researchers have used self-report inventories to assess Gardner's multiple intelligences (e.g., Shearer, 1996; Teele, 1992). However, previous research has suggested that self-report measures of intelligence correlate only modestly with maximum performance measures (Borkenau & Liebler, 1993; Furnham & Fong, 2000). Paulhus, Lysy, and Yik (1998) suggested that even when aggregated measures were used, the validity limit of self-estimates of intelligence was 0.30.

3. The intelligences

Below, the eight "intelligences" as conceptualized by Gardner (1993) are described in detail, with the aim of identifying the range of abilities subsumed by each domain and of examining the cognitive demands of tasks assessing these abilities. In describing the abilities within the domains, we make frequent reference to previous research on the structure of cognitive abilities. In particular, we draw comparisons with the "primary mental abilities" or group factors that were first identified by Thurstone (1938), and also with the factors at all three levels (or strata) of the hierarchy described by Carroll (1993). Based on this analysis, the broad categories of tasks to be assessed as indicators of the various ability domains will be identified.

3.1. Linguistic intelligence

Gardner has described Linguistic intelligence as a sensitivity to spoken and written language and the ability to use language to accomplish goals, as well as the ability to learn new languages. Lawyers, public speakers, writers, and poets all possess high levels of Linguistic intelligence, according to Gardner.

The Linguistic intelligence domain, as described by Gardner, seems to encompass a wide variety of more specific abilities. Thurstone (1938), for example, differentiated between verbal comprehension and word fluency, which represented two of his seven primary mental abilities, whereas Gardner would include both under the domain of Linguistic intelligence. Verbal comprehension involves the ability to understand the meanings both of individual words and of passages of written or spoken text. Word fluency, in contrast, involves the ability to generate rapidly many examples of words that meet some specification (e.g., words beginning with a given letter, words rhyming with a target word, words naming objects that have some property, etc.). Previous studies have suggested that indicators of verbal comprehension tend to be slightly more highly *g*-loaded than are measures of word fluency (e.g., median *g*-loadings of 0.49 versus 0.43 reported by Carroll, 1993, p. 597).

3.2. Spatial intelligence

Gardner defined Spatial intelligence as the ability to recognize both large and small visual patterns. He suggested that navigators and pilots would possess high levels of spatial intelligence, as would sculptors, surgeons, chess players, and architects.

Previous research in the domain of spatial abilities suggests that spatial visualization and spatial scanning are two important and distinct aspects of that domain (e.g., Ekstrom, French, Harman, & Derman, 1976). Spatial visualization refers to the ability to imagine the movement of an object and is typically measured with mental rotation tasks. Carroll (1993) noted that visualization tasks generally form a first-stratum factor, and one that tends to be highly g-loaded. Spatial scanning is the ability to scan a field quickly, to follow paths visually, and to reject false leads (Ekstrom et al., 1976). Carroll (1993) tentatively identified this capacity as a first-order factor, but stated that further research was necessary before it could be considered independent and interpreted accordingly. Tasks assessing spatial visualization and spatial scanning tend to load on a second-stratum factor of broad visualization ability, which corresponds also to Thurstone's (1938) spatial ability factor.

3.3. Logical/Mathematical intelligence

Gardner described Logical/Mathematical intelligence as the ability to study problems, to carry out mathematical operations logically and analytically, and to conduct scientific investigations. Gardner identified mathematicians, logicians, and scientists as persons who would possess high levels of this hypothesized intelligence.

Reasoning, the domain whose content is subsumed within the definition of Gardner's Logical/Mathematical intelligence, was identified as one of the primary mental abilities recovered by Thurstone (1938). According to Carroll (1993), reasoning subsumes six first-stratum factors: general reasoning, verbal reasoning, induction, quantitative reasoning, syllogistic reasoning, and classification ability. Quantitative reasoning, which combines numerical content with logical thinking, would seem to be a prototypical exemplar of Gardner's Logical/Mathematical intelligence domain. Carroll (1993) found that the first-stratum factor of quantitative reasoning was highly g-loaded, as were other reasoning abilities, such as induction.

The Logical/Mathematical domain of Gardner's framework would also subsume numerical facility, which is measured with tasks requiring participants to quickly perform simple arithmetic computations, such as addition, subtraction and multiplication. This numerical skill emerged as one of the primary mental abilities in Thurstone's (1938) research, defining a different factor from that which subsumed reasoning tasks, although *quantitative* reasoning also shows some association with this factor. In Carroll's (1993, p. 597) review, a first-stratum factor of numerical facility was somewhat less g-loaded than was that of quantitative reasoning.

3.4. Interpersonal intelligence

According to Gardner (1983), an individual who is high in Interpersonal intelligence understands the intentions, motivations, needs, and desires of others, and is capable of working effectively with them. Gardner stated that teachers, clinicians, salespeople, politicians, and religious leaders all use Interpersonal intelligence.

Gardner's Interpersonal intelligence would seem to be related to the construct of emotional intelligence, which can be associated with intelligence or with personality depending on how it is measured. For example, O'Connor and Little (2003) reported that an ability-based measure of emotional intelligence was correlated more strongly with cognitive ability than with personality. A self-report inventory of emotional intelligence, on the other hand, was correlated more strongly with personality than with cognitive ability.

The Interpersonal domain would seem to include both an understanding of verbal and nonverbal social cues. The individual with a high level of Interpersonal ability would likely possess both an awareness of the social consequences of events and also an understanding of the motivations and intentions underlying people's behavior. Thus, this domain could be assessed by asking an individual to anticipate the development of a social situation, or to infer the state of mind of a person based on his or her words or actions.

3.5. Intrapersonal intelligence

Gardner (1999) described Intrapersonal intelligence as the ability to understand and to have an effective working model of oneself. Intrapersonal intelligence, as conceptualized by Gardner, includes the awareness of one's own desires, fears, and abilities, and also using this information to make sound life decisions.

From Gardner's description, it appears that having a clear concept of oneself is a key component of his Intrapersonal domain. In previous research, self-concept clarity was operationalized in an investigation of the nature of self-esteem (Campbell, 1990) in which participants made "me/not me" decisions for a 56-item list of adjectives, within which were 25 pairs of opposite poles of various personality traits. Campbell was then able to examine the inconsistency of participants' self-descriptions by determining to what extent they endorsed opposites to describe themselves. Results indicated that this measure effectively distinguished high self-esteem and low self-esteem groups, which were hypothesized to differ in self-concept clarity. Therefore, assessments of self-concept clarity might serve as an indicator of Intrapersonal ability.

Intrapersonal intelligence, as described by Gardner, is also somewhat related to metacognition in general and to the ability to self-monitor in particular. That is, the individuals with high Intrapersonal ability should be aware of what they know as well as what they do not know. However, Stankov (2000) reported that his research has found very little correlation between self-monitoring ability, as measured by the difference between a confidence score and the actual percentage of correctly solved items, and intelligence. These findings could be interpreted as support for Gardner's contention that Intrapersonal ability is an independent area of intelligence. Thus, measures of the extent to which individuals can accurately judge their relative strengths and weaknesses might serve as an index of Intrapersonal ability.

3.6. Naturalistic intelligence

Gardner (1999) described a naturalist as one who is able to recognize and classify objects. According to Gardner, hunters, farmers, and gardeners would have high levels of Naturalistic intelligence, as would artists, poets, and social scientists, who are also adept at pattern-recognition. He stated that a marketing professional who promotes the small differences between competing products is applying Naturalistic intelligence, as is the individual who can recognize cars from the sounds of their engines.

As described above, a central element of Gardner's Naturalistic intelligence is the capacity to categorize objects according to salient similarities and differences among them. This ability is critically involved in the generation of meaningful taxonomies of both living and non-living objects. Therefore, categorization tasks of this kind would appear to be ideal measures of the Naturalistic domain. It is worth noting that these tasks also appear to demand a high level of logical reasoning, which suggests that cognitive demands for this domain might in fact be similar to those for Gardner's Logical/Mathematical intelligence, despite being applied to the realm of semantically meaningful stimuli rather than to the domain of symbolic, quantitative concepts.

3.7. Bodily-Kinesthetic intelligence

Gardner (1999) described this intelligence as the potential of using the whole body or parts of the body in problem-solving or the creation of products. Gardner identified not only dancers, actors, and athletes as those who excel in Bodily-Kinesthetic intelligence, but also craftspeople, surgeons, mechanics, and other technicians. Thus, Gardner does not appear to differentiate between gross motor skills (i.e., involving the whole body or the larger muscle groups) and fine motor skills (i.e., involving smaller muscle groups, especially those controlling the hands and fingers) in describing Bodily-Kinesthetic intelligence. Gardner has not explained why these abilities would be expected to be strongly associated with each other. Given that the Bodily-Kinesthetic domain subsumes both gross and fine motor skills, the assessment of this domain would require measurements of both of these intuitively rather distinct areas of ability.

3.8. Musical intelligence

Gardner (1999) suggested that Musical intelligence was parallel in structure to Linguistic intelligence, and that it would be reflected in the performance, composition, and appreciation of musical patterns. With regard to the underlying abilities involved in his Musical intelligence, Gardner has claimed that the two most central constituent elements of music are rhythm and pitch (or melody), followed in importance by timbre (which Gardner, 1983, p. 105, describes as the characteristic qualities of a tone). This is consistent with the content of eight first-stratum factors identified by Carroll (1993) as aspects of a broader Auditory Receptive Ability (defined by 12 first-stratum factors in all). The eight music-relevant factors included the following: discrimination of tones and sequences of tones with respect to basic attributes such as pitch, intensity, duration, and rhythm; auditory cognitive relations (judgments of complex relations among tonal patterns); tonal imagery; discrimination and judgment of tonal patterns in musicality; temporal tracking; ability to recognize and maintain mentally an equal-time beat; ability to retain, on a short-term basis, images of tones, tonal patterns, and voices; and absolute pitch ability. Thus, given that rhythm and tone would appear to be core aspects of these narrow factors of musical ability, measures of the abilities to discriminate between rhythms and between tones would be important elements in the assessment of Gardner's Musical intelligence.

Some previous research has examined the extent to which measures of musical ability are loaded on the *g* factor. For example, Lynn, Wilson, and Gault (1989) found positive correlations between musical ability and general intelligence ranging from 0.27 to 0.40 in a sample of 10-year-olds, suggesting that musical ability is not independent of *g*.

4. The current study

The purpose of the current study was to investigate Gardner's Multiple Intelligences theory by examining the relations among the eight hypothesized intelligences, as well as their relations with an external measure of general cognitive ability. 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. Gardner (1999) cautioned that there is a tendency to measure non-verbal abilities with verbal measures, leading to artificially high correlations among the ability domains. We therefore avoided measuring variance attributable to verbal ability by including several nonverbal measures of the various hypothesized intelligences. For example, Logical-Mathematical, Spatial, Interpersonal, Musical, and Bodily-Kinesthetic intelligences were tested with at least one of the two measures being entirely nonverbal. In all other tests, with the exception of Linguistic, reading demands were minimal, so verbal ability could not be considered a source of shared variance between these tests.

In generating hypotheses for the current study, we first made a distinction between two categories of "intelligences" of Gardner's model, according to the extent to which non-cognitive abilities were involved in each domain. Based on the descriptions of the domains (see summaries above), Gardner's Linguistic, Spatial, Logical-Mathematical, Naturalistic, and Interpersonal intelligences would appear to be purely cognitive in nature. Intrapersonal ability, on the other hand, would seem to have some element of personality in addition to its cognitive component. (Interpersonal might also have this personality aspect when applied in real-life settings.) Gardner's Bodily-Kinesthetic intelligence chiefly involves motor rather than cognitive abilities, and his Musical intelligence clearly has a sensory as well as a cognitive component.

Our first hypothesis involved the extent to which measures of the various intelligences would be correlated, and hence the extent to which they would define a *g* factor of general intelligence. Given the recurrent findings of positive manifolds in the matrices of correlations among tests of cognitive ability (e.g., Spearman, 1927), and the corresponding strong empirical support for a *g* factor of general intelligence, we hypothesized similar findings for the present study. More precisely, using the above-noted distinction between purely cognitive and partly non-cognitive abilities, we expected that all of the tests except those in the partly non-cognitive domains would be correlated with each other and with an independent, external measure of general intelligence. We also predicted that a factor analysis of the tests would yield a *g* factor, with all of the purely cognitive tests having substantial *g*-loadings, and with all of the remaining tests having lower *g*-loadings. In addition, we expected that the pattern of tests' *g*-loadings would also be reflected in the tests' correlations with an external marker of *g*, specifically, a brief test of general intelligence.

Our second hypothesis involved the degree of coherence of each of the intelligence domains, or the extent to which the distinct abilities assessed by our measures of each domain would in fact be correlated, beyond the influence of *g*. Given the recurrent empirical findings of group factors broadly similar to several of Gardner's intelligences, we hypothesized that the two tests within each intelligence domain (with one exception) would be correlated above and beyond their association with the general factor. In other words, we expected that residual correlations between the tests assessing the same domain, after extraction of the *g* factor, would tend to be positive. For the Bodily-Kinesthetic domain, however, we expected that the measures of gross- and fine-motor skills would be nearly uncorrelated, given the apparent dissimilarity of those abilities.

We should mention that we chose not to specify a confirmatory factor-analytic (CFA) model in testing Gardner's theory. This decision was based in part on the small number of tests per domain, as testing time constraints limited us to only two tests for each of the eight hypothesized intelligences; a properly specified CFA model would require at least three, and ideally more, indicators of each factor. In addition, given the inherent factorial complexity and lack of simple structure associated with individual difference variables, it seemed likely that–even if Gardner's theory were largely valid–the results of CFA evaluations would reject his model, in the same way that even very widely replicated models of personality structure are also rejected by CFA for the same reasons (see McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). Thus, to avoid these problems of underdefined factors and of unrealistically stringent tests of Gardner's model, we instead focused our analyses on the tests' intercorrelations, *g*-loadings, correlations with an external measure of *g*, and residual intercorrelations after extraction of *g*.

5. Method

5.1. Participants

Two hundred participants (116 women) were recruited through posters, in-class presentations, and sign-up sheets on two university campuses. To recruit participants, presentations were made in a variety of classes, including chemistry, accounting, sports management, psychology, and geography. Participants were required to speak English as their first or best language and to be at least 18 years of age. Of the 200 participants, 171 were current undergraduate students; the remainder included current graduate students, non-academic employees of the university, and non-student friends or relatives of the undergraduate student participants. Participants ranged in age from 17 to 66 years (M=22.7, SD=6.1). For the 186 participants who had attended university, the academic major fields represented 30 different academic departments, with self-reported academic average grades in university courses ranging from 56% to 94% (M=75.9, SD=7.6) for the most recent academic year.

5.2. Measures

5.2.1. Demographic information

Participants provided the following demographic information: age, sex, faculty in which the student is enrolled, subject in which the student is majoring, year of study, and previous year's average mark, expressed as a percentage.

5.2.2. General cognitive ability

The Wonderlic Personnel Test (WPT; Wonderlic, 2002) Form A was used as a measure of general intelligence external to the tests representing Gardner's intelligence domains. The WPT is an extensively normed 50-item, 12-min omnibus test. WPT items represent a broad range of problem types (e.g., analogies, arithmetic problems, proverbs, geometric figures, number series, and word definitions) and are presented in order of increasing difficulty.

5.2.3. Linguistic intelligence

The Educational Testing Service (ETS) Advanced Vocabulary Test II is a 36-item, 8-min test of verbal comprehension, appropriate for a university undergraduate population. For each target word, participants select one of four words which has the same or nearly the same meaning as the target word.

The ETS Opposites Test assesses associational fluency by requiring participants to generate words that are opposite or nearly opposite in meaning to a given word, for each of eight target words (four on each 5-min half of the test).

5.2.4. Spatial intelligence

The ETS Map Planning Test is a 40-item, 6-min test of spatial scanning. For each of two parts of the test, participants are presented with a map showing streets, roadblocks, and numbered buildings. Participants must find the shortest route between two lettered points, avoiding roadblocks. The correct response is the number of the building which is passed in traveling the shortest route between two given points.

The ETS Paper Folding Test is a 20-item, 6-min measure of the ability to make serial mental rotations of spatial stimuli, which Carroll (1993) identified as a visualization task. Each item consists of drawings of a square of paper that is folded two or three times, in various ways, and then has a hole punched through it. Participants choose which of five drawings accurately depicts how the sheet of paper and its pattern of holes would appear when the paper is fully opened.

5.2.5. Logical/Mathematical intelligence

The ETS Subtraction and Multiplication Test is a 120-item, 4-min measure of the ability to perform basic arithmetic operations with speed and accuracy.

The ETS Necessary Arithmetic Operations Test is a 30-item, 10-min measure of general arithmetic reasoning. The problems require the participant to determine which arithmetic operations are required to solve mathematical word problems, but do not require the participant to perform those computations.

5.2.6. Interpersonal intelligence

Cartoon Predictions (O'Sullivan & Guilford, 1965a) is a 10-min test with 30 items, each consisting of one cartoon showing people's reactions in a situation involving two or more people. Participants must choose one of three related cartoons to indicate what would most logically happen next. For example, the sample item shows a man clinging to the roof of the house, while a boy looks on. The man appears to be frightened and asking for help. The correct response shows the boy with a woman bringing a ladder to help the man down. The two incorrect responses show the man trying to climb on the roof, and the boy and the woman laughing at the man. The score is the number of correct responses plus $1/k$ times the number of items left blank, k being the number of alternative answers. Thus, this test measures an ability to anticipate the sequence of events in a social situation, without drawing on verbal ability.

Social Translations (O'Sullivan & Guilford, 1965b) is a 10-min test with 24 items. Each item lists a brief statement along with the identities of the person by whom and to whom it was said. Participants choose one of three pairs of people for which the same verbal statement would have a meaning different from the meaning for the original pair. For example, the phrase "That's mine" said by a man to a parking lot attendant is similar in meaning when said by a student to a librarian or by a busy shopper to a salesclerk, but is different in meaning when said by an angry child to a playmate. This test thus assesses an ability to understand differences in the motivations of persons who make the same statement, according to the social context in which the statement is made.

5.2.7. Intrapersonal intelligenc

One Intrapersonal measure involved the consistency of self-ratings on personality-descriptive adjectives that define the same broad trait. Participants completed a personality self-rating form in which they rated, on a 9-point scale, how closely each of 48 adjectives described their typical personality. (The adjectives were selected by the second author as markers of the six personality factors derived from the personality lexicon of the English language; see Ashton, Lee, & Goldberg, 2004) The standard deviation of scores was calculated for each individual across his or her self-ratings on adjectives within each factor (after recoding of reverse-keyed items). The average standard deviation across factors, each of which can range from 0 to 4, indicates inconsistency of self-description; this average value, when multiplied by −1, was used as an index of consistency of self-description, which represents self-concept clarity as discussed in the Introduction.

The second Intrapersonal measure was an accuracy coefficient, which reflected how closely participants' self-ratings in each ability domain matched their actual performance. As previously described, participants rated their ability on each of nine domains corresponding to Gardner's eight intelligences as well as overall intelligence. For each ability, participants indicated which percentage of their fellow students would have a lower level of that ability than themselves. A within-person correlation between self-estimated ability and actual ability was calculated across the eight abilities after conversion of percentiles to standard scores. This value represents the self-monitoring or self-assessment ability described in the Introduction, by indicating the individual's understanding of his or her own pattern of relative strengths and weaknesses.

5.2.8. Naturalistic intelligence

ETS Making Groups is a 4-item, 10-min test of the ability to categorize objects into logical groups. Each item consists of a list of seven objects. The participant is asked to produce up to 10 groups of three or more objects, and to provide a very brief reason for each grouping. For example, a respondent might classify airplanes, boats, and cars together as "vehicles", classify bats, robins, and frogs together as "animals", and classify airplanes, bats, and robins together as "things that fly".

ETS Diagramming Relationships is a 30-item, 8-min test of the comprehension of the relations among various categories of objects. For each item, a group of three categories of objects is given, and participants are asked to select one of five Venn diagrams that best illustrates the relationships among those categories of objects. For example, the categories dogs, cats, and animals would be best depicted with two non-overlapping circles (representing cats and dogs) completely inside a larger circle (representing animals).

5.2.9. Bodily-Kinesthetic intelligence

A test of balance was chosen to represent gross motor ability, because balance is a key element in assessing potential in gymnastics, diving, skating, and many other athletic activities (Johnson & Nelson, 1986). In addition, balance seemed to be an ability that would be less susceptible to training effects than would many other potential abilities, such as strength, endurance, or speed. Specifically, we used the Stork Stand (Johnson & Nelson, 1986) as the test of balance. Participants were timed while standing with hands on hips, one leg raised and bent with its foot on the inside of the other knee. On a signal, participants raised their heels so that only the ball of one foot remained on the floor. Each participant's score was the average of the two longest times in seconds between raising the heel and losing balance, from three trials.

The General Aptitude Test Battery (GATB; U.S. Department of Labor, 1970) Mark Making test, a measure of manual dexterity, was used to measure fine motor ability. Participants were required to make marks consisting of two parallel lines with another line underscoring the two. Participants were asked to make one such mark in each of many squares on the printed page, filling as many squares as possible in 60 s.

5.2.10. Musical intelligence

GIA Publications' Advanced Measures of Music Audiation (AMMA; Gordon, 1989) is a 30-item, 16-min test, consisting of sets of musical statements played to participants on a CD player. Three practice items are included. For each set, the participant must decide whether the two musical statements are the same, tonally different, or rhythmically different.

No musical background is required in order to take the AMMA. The test is based on Gordon's (1989) contention that audiation–the ability to hear mentally and to comprehend musical patterns when the sound is no longer present–is the basis of musical aptitude. The assumption underlying the AMMA is that being able to hold musical patterns in memory and to identify the presence and type of discrepancy between them is a better measure of musical aptitude than is musical achievement, which presumably depends in part on musical training. Each item consists of an equal number of pairs of musical patterns, such that the patterns may or may not differ in their tone or their rhythm. Specifically, 10 items have patterns that differ in tone, that is, in individual pitch, in mode tonal center (the musical key of the melody), or in a combination of these. Similarly, 10 items have patterns that differ in rhythm, that is, in durations, meters, tempos, or combinations of these. Finally, 10 other items have patterns that are identical.

Separate scores are calculated for the Tonal and Rhythm scales. The Tonal score (40 points possible) is calculated by summing the number of correct tonal responses (10 points possible) and the number of correct identifications of sameness (10 points possible) plus 20 points. The number of erroneous tonal identifications is then subtracted from the score, with omitted items counted neither for nor against. Rhythm (40 points possible) is calculated similarly. For the purposes of all analyses in the current study, however, only the first half of the Rhythm test and the second half of the Tonal test were used, with scores potentially ranging from 0 to 20. This approach was taken to correct for the fact that Tonal and Rhythm total scores would otherwise show a strong artifactual correlation due to overlapping content in incorrect responses. But, in order to calculate the reliability of the Rhythm and Tonal tests, we correlated each of the half-length tests used in the analysis with its other half (i.e., the half that was not used in the main analyses below).

5.3. Procedure

Participants who met the inclusion criteria (183years or older, English as first or best language) were tested in groups of 2–36 in various meeting rooms on a university campus. All participants were given information letters as well as an oral description of procedures prior to providing signed informed consent. The testing session took place over a 2.5–3 h period, with a 10-min break at approximately the halfway point, following completion of the music test.

The 15 measures were administered in two different orders. The first order was as follows: demographic information; Opposites (Linguistic); Map Planning (Spatial); Subtraction and Multiplication (Logical/Mathematical); Cartoon Predictions (Interpersonal); Diagramming Relationships (Naturalistic); Stork Stand (Bodily-Kinesthetic); AMMA (Rhythm and Tonal; Musical); 10-min break; Making Groups (Naturalistic); Social Translations (Interpersonal); Necessary Arithmetic Operations (Logical/Mathematical); Paper Folding (Spatial); personality adjective self-ratings (Intrapersonal); and Mark Making (Bodily-Kinesthetic). The second order reversed the presentation of the Linguistic, Spatial, Logical/Mathematical, Interpersonal, and Naturalistic tests. The second order was as follows: demographic information; Making Groups (Naturalistic); Social Translations (Interpersonal); Necessary Arithmetic Operations (Logical/Mathematical); Paper Folding (Spatial); Stork Stand (Bodily-Kinesthetic); AMMA (Rhythm and Tonal; Musical); 10-min break; Opposites (Linguistic); Map Planning (Spatial); Subtraction and Multiplication (Logical/Mathematical); Cartoon Predictions (Interpersonal); Diagramming Relationships (Naturalistic); personality adjective self-ratings (Intrapersonal); and Mark Making (Bodily-Kinesthetic). One hundred and one (50.5%) participants completed tests in the first order and the remainder completed them in the second.

Each participant was paid $30 upon completion of the testing session and provided with a feedback letter that briefly explained the purpose of the research.

6. Results

6.1. Measured multiple intelligences

6.1.1. Reliabilities

The reliabilities of the ability tests are given in Table 1. Note that the value for the WPT is coefficient alpha, whereas all others are split-half reliabilities; no reliability value is reported for Accuracy of self-ratings (an Intrapersonal test), because it did not include multiple items from which internal consistency or split-half reliability could be calculated. The extremely low reliability of both music tests was surprising, given the high split-half reliabilities reported in the AMMA manual (specifically, values of 0.80 for both Tonal and Rhythm in undergraduate and graduate non-music majors). The reliabilities of Cartoon Predictions (an Interpersonal test) and of Consistency (an Intrapersonal test) were also low. All remaining reliability coefficients were 0.67 or greater, with almost half of the ability tests having coefficients of 0.80 or greater.

6.1.2. Descriptive statistics

Minimum and maximum scores, means, and standard deviations of the tests measuring the eight intelligence

Table 1
Descriptive statistics and split-half reliabilities of measures of each intelligence domain and *g*

Measure	Intelligence domain	Minimum	Maximum	Mean	SD	Reliability
Wonderlic	*g*	9.00	44.00	26.60	5.87	0.80
Opposites	Linguistic	9.00	40.00	23.98	5.76	0.67
Vocabulary	Linguistic	5.00	33.50	19.36	5.32	0.67
Map planning	Spatial	1.00	39.00	22.90	7.40	0.80
Paper folding	Spatial	0.26	20.00	10.50	4.00	0.76
Subtraction and Multiplication	Logical/Mathematical	5.00	115.00	41.60	15.82	0.94
Necessary arithmetic operations	Logical/Mathematical	0.34	28.68	14.32	5.92	0.81
Cartoon predictions	Interpersonal	10.00	28.00	20.86	3.02	0.45
Social translations	Interpersonal	4.00	23.00	17.36	3.86	0.90
Accuracy	Intrapersonal	−0.75	0.93	0.23	0.41	N/A
Consistency	Intrapersonal	−3.28	−0.91	−1.77	0.38	0.53
Diagramming Relationships	Naturalistic	0.00	30.00	16.92	6.94	0.82
Making groups	Naturalistic	5.00	29.00	18.86	4.58	0.79
Stork stand	Bodily-Kinesthetic	0.00	120.00	32.74	31.27	0.92
Mark making	Bodily-Kinesthetic	46.00	120.00	77.84	10.52	0.93
Rhythm	Musical	8.00	19.00	13.79	2.31	0.28
Tonal	Musical	7.00	17.00	11.78	2.39	0.17

SD = Standard Deviation.
N = 200.

domains are given in Table 1. We also calculated descriptive statistics for the external measure of general intelligence, the WPT, in order to compare the overall level of cognitive ability in our sample with that of the population at large. In this sample, the mean and standard deviation of raw scores on the WPT were 26.60 and 5.87, respectively, which compare with values of 21.06 and 7.12 for the WPT general population normative sample (Wonderlic, 2002, inside front cover). Thus, the mean level of general intelligence was considerably higher than that of the general population, and equivalent to an IQ score of 114 (see Wonderlic, 2002, p. 20). As would be expected given the composition of our predominantly university student sample, the distribution of WPT scores was similar to those reported for male (M=26.51, SD=6.03) and female (M=24.95, SD=5.59) subsamples of the college population normative sample (Wonderlic, 2002, p. 38).

The standard deviation of WPT scores in our sample was about 82% as large as that of the WPT normative sample, which indicates some restriction of variance in ability relative to that of the general population. As a result, the correlations among the tests will tend to be smaller than would be observed in a general population sample. Note, however, that despite the restricted variation in the present sample, the range in WPT scores was still rather wide: three participants had IQ-equivalent WPT scores below 85, and three others had 140 or above.

Mean scores on the ability tests were generally similar for the male and female participants of the present sample. Tests of the differences between men and women reached statistical significance ($p<0.05$) for only three subtests, and even for those tests the sizes of those differences were modest: women outperformed men in the Making Groups and Cartoon Predictions subtests ($d=0.36$ and $d=0.41$, respectively), and men outperformed women in Necessary Arithmetic Operations ($d=-0.29$).

6.1.3. Test intercorrelations

Correlations among the tests measuring the intelligence domains are shown in Table 2. As hypothesized, the correlation matrix showed a positive manifold. Moreover, when the tests of the purely cognitive abilities are considered, 34 of the 45 (75.6%) correlations among those tests reached the conventional level of statistical significance. The smallest of the correlations among the purely cognitive tests were those (a) involving the Subtraction and Multiplication test (a numerical facility test representing Logical/Mathematical intelligence), and (b) between tests of Linguistic and Spatial abilities. Nevertheless, all of those subtests showed substantial correlations with the Necessary Arithmetic Operations test (a quantitative reasoning test representing the Logical/Mathematical domain) and with the tests of Naturalistic ability (particularly the Diagramming Relationships test). This latter result suggests that a common element of ability, related to reasoning, underlies performance across the purely cognitive tests.

Table 2
Correlations among ability tests

Test	Vocab.	Map Plan	Paper Fold	Subt/Mult	Nec. Arith	Cart. Pred	Social Tran	Accur.	Cons.	Diag. Rel.	Mak. Grps.	Stork Stand	Mark Mak	Rhythm	Tonal
Opp.	**0.55****	0.10	0.07	0.07	0.32**	0.22**	0.24**	0.18**	0.10	0.38**	0.40**	0.01	−0.05	0.03	0.06
Vocab		0.10	0.08	0.10	0.36**	0.32**	0.39**	0.18*	0.17*	0.37**	0.27**	−0.05	−0.08	0.18*	0.03
Map Plan			**0.46****	0.26**	0.46**	0.09	0.23**	0.05	0.21**	0.44**	0.40**	0.04	0.17*	0.04	0.10
Paper Fold				0.09	0.45**	0.20**	0.23**	0.00	0.10	0.43**	0.32**	0.06	0.01	0.11	−0.02
Subt/Mult					**0.32****	−0.09	0.09	−0.05	0.12	0.18**	0.06	−0.02	0.18*	−0.01	0.05
Nec. Arith.						0.23**	0.31**	−0.02	0.18*	0.55**	0.32**	0.02	−0.07	0.14	0.06
Cart. Pred.							**0.28****	0.23**	0.11	0.22**	0.18*	0.00	−0.01	0.07	0.02
Soc. Tran.								0.22**	0.19**	0.39**	0.28**	−0.09	0.02	0.14	0.04
Accuracy									**0.01**	0.05	0.11	0.09	0.11	0.00	−0.06
Consist.										0.15*	0.19**	0.05	−0.06	−0.02	−0.03
Diag. Rel.											**0.42****	0.04	0.03	0.15*	0.12
Mak. Grps												0.05	0.10	0.04	0.04
Stork Stand													**0.13**	0.18*	0.16*
Mark Mak.														0.06	0.14*
Rhythm															**0.19****

Opp. = Opposites; Vocab. = Vocabulary; Map Plan. = Map Planning; Paper Fold. = Paper Folding; Subt/Mult = Subtraction and Multiplication; Nec. Arith. = Necessary Arithmetic Operations; Cart. Pred. = Cartoon Predictions; Soc. Tran. = Social Translations; Consist. = Consistency; Diag. Rel. = Diagramming Relationships; Mak. Grps = Making Groups; Mark Mak. = Mark Making. Correlations in bold type involve tasks within the same domain. N = 200. *p < 0.05. **p < 0.01, two-tailed.

In contrast to the substantial correlations observed between most pairs of purely cognitive tests, only 18 of the 75 (24%) correlations in which at least one of the variables was partly non-cognitive in nature reached statistical significance. In particular, the correlations involving the Musical or the Bodily-Kinesthetic tests were almost all very small; recall, however, that both of the Musical tests had extremely low reliabilities.

6.3. Test g-loadings and correlations with WPT

We examined the *g*-loadings of the entire set of ability measures representing the eight hypothesized intelligences by extracting the first unrotated factor from those 16 tests, using principal axis factoring. Eigenvalues for the first seven factors were 3.74 (23.4% of total variance), 1.64 (10.3%), 1.41 (8.8%), 1.19 (7.4%), 1.13 (7.0%), 0.96 (6.0%), and 0.90 (5.6%). A clear break in eigenvalue sizes was observed between the second and first factors only, thus suggesting only one large factor in this data set. When we extracted that single, general factor, the tests showed a wide range of *g*-loadings (see Table 3, which also provides loadings corrected for the imperfect reliability of the tests). Two tests–Necessary Arithmetic Operations (Logical/Mathematical) and Diagramming Relationships (Naturalistic)–had very high *g*-loadings, exceeding 0.70. Several tests showed fairly high *g*-loadings, between 0.40 and 0.59, including Opposites (Linguistic), Paper Folding (Spatial), Social Translations (Interpersonal), Vocabulary (Linguistic), Map Planning (Spatial), and Making Groups (Naturalistic). Modest *g*-loadings, between 0.20 and 0.39, were observed for Subtraction and Multiplication (Logical/Mathematical), Consistency (Intrapersonal), and Cartoon Predictions (Interpersonal). The two Bodily-Kinesthetic tests–Stork Stand and Mark Making–had *g*-loadings under 0.10, and three other tests–Tonal (Musical), Accuracy (Intrapersonal), and Rhythm (Musical)–also had *g*-loadings below 0.20. The overall pattern of *g*-loadings suggests that the purely cognitive tests, except Subtraction and Multiplication, were strongly saturated with a general cognitive ability, and that the reasoning-based tests (Necessary Arithmetic Operations and Diagramming Relationships) were particularly strong indicators of *g*.

(Note that the relatively modest size of the *g* factor as computed above was largely attributable to the inclusion of tests with very little cognitive content. When tests from the Bodily-Kinesthetic, Intrapersonal, and Musical domains were removed from the factor analysis, the variance accounted for by *g* increased to 35.8%.)

As an alternative means of assessing the tests' saturation with general intelligence, we computed the

Table 3
g loadings of tests and correlations of tests with Wonderlic Personnel Test (WPT)

Ability Domain	Test	g-loading	$r_{(\mathrm{WPT})}$
Linguistic	Opposites	0.50 (0.61)	0.41** (0.56)
	Vocabulary	0.54 (0.66)	0.47** (0.64)
Spatial	Map Planning	0.55 (0.61)	0.48** (0.60)
	Paper Folding	0.50 (0.57)	0.48** (0.62)
Logical/ Mathematical	Subtraction and Multiplication	0.24 (0.25)	0.36** (0.42)
	Necessary Arithmetic Operations	0.70 (0.78)	0.67** (0.83)
Interpersonal	Cartoon Predictions	0.37 (0.55)	0.23** (0.38)
	Social Translations	0.53 (0.56)	0.38** (0.45)
Intrapersonal	Accuracy	0.16 (N/A)	0.11 (N/A)
	Consistency	0.27 (0.37)	0.27** (0.41)
Naturalistic	Diagramming Relationships	0.75 (0.83)	0.59** (0.73)
	Making Groups	0.57 (0.64)	0.38** (0.48)
Bodily- Kinesthetic	Stork Stand	0.03 (0.03)	−0.04 (−0.05)
	Mark Making	0.06 (0.06)	0.03 (0.03)
Musical	Rhythm	0.18 (0.34)	0.08 (0.17)
	Tonal	0.10 (0.24)	0.07 (0.19)

Values in parentheses are corrected for unreliability in the individual ability tests only (for the g-loadings) or in both the individual ability tests and the WPT (for the WPT correlations).
$N=200$. *$p<0.05$. **$p<0.01$, two-tailed.

correlations of those tests with the WPT (see Table 3, which also provides correlations corrected for the imperfect reliability of the ability tests and of the WPT). Consistent with our hypotheses, all of the purely cognitive measures (i.e., Linguistic, Spatial, Logical/Mathematical, Interpersonal, and Naturalistic tests) were significantly correlated with that measure of general intelligence. Only the Bodily-Kinesthetic and Musical tests, as well as one of the Intrapersonal tests (Accuracy), failed to show statistically significant positive correlations with the WPT. Necessary Arithmetic Operations and Diagramming Relationships had the greatest correlations with the WPT, a result that is consistent with the high g-loadings of those tests in the previous analysis.

6.3.1. Coherence of abilities within domains, beyond g

The coherence of the tests within each intelligence domain was examined first through an inspection of zero-order correlations within and between domains (Table 2), and then through an inspection of residual correlations, after removal of the g factor of Table 3, between tests in the same domain (see Table 4). With regard to the within-domain zero-order correlations, the correlation between the two Linguistic tests was the highest involving either test. This was also the case for the Musical tests. For the Spatial tests, Map Planning and Paper Folding were substantially intercorrelated, but both tests also showed similarly strong relations with Necessary Arithmetic Operations (Logical/Mathematical) and Diagramming Relationships (Naturalistic). In the Logical/Mathematical domain, Subtraction and Multiplication had its highest correlation with Necessary Arithmetic Operations, but the latter test showed similar or higher correlations with seven other tests. For the Naturalistic tests, Making Groups had its highest correlation with Diagramming Relationships, but the latter test was more highly correlated with three other tests from the Spatial and Logical/Mathematical domains. The two Interpersonal tests, Social Translations and Cartoon Predictions, were significantly correlated with each other, but each had higher correlations with other tests. Only for the Intrapersonal and Bodily-Kinesthetic domains did the within-domain correlations fail to reach the conventional level of statistical significance.

In addition to examining the zero-order correlations between tests within an intelligence domain, we also considered the residual correlations following extraction of the g factor. This allows an evaluation of the extent to which measures within each domain are related beyond the influence of g. The residual correlations (see Table 4) ranged from −0.04 (Intrapersonal) to 0.29 (Linguistic). As indicated by the near-zero residual correlations between the two Intrapersonal tests, between the two Naturalistic tests, and between the two Interpersonal tests, the indicators of each of these domains were essentially unrelated to each other after the removal of g variance; this result indicates that the substantial zero-

Table 4
Residual correlations within each intelligence domain after extraction of g factor

Intelligence Domain (Tests)	Residual Correlation
Linguistic (Opposites, Vocabulary)	0.29**
Logical/Mathematical (Necessary Arithmetic Operations, Subtraction and Multiplication)	0.15*
Spatial (Map Planning, Paper Folding)	0.18*
Interpersonal (Cartoon Predictions, Social Translations)	0.09
Intrapersonal (Accuracy, Consistency)	−0.04
Naturalistic (Making Groups, Diagramming Relationships)	−0.01
Bodily-Kinesthetic (Stork Stand, Mark Making)	0.13
Musical (Rhythm, Tonal)	0.17*

$N=200$.
*$p<0.05$. **$p<0.01$, two-tailed.

order correlations within the latter two domains were attributable to *g*, and not to any domain-specific similarity. The small residual correlations between the two Bodily-Kinesthetic tests and between the two Musical tests were nearly identical in size to the corresponding zero-order correlations, a result which suggests that the little variance shared by the tests within these domains is not attributable to *g*, but rather to variance specific to each of these "intelligences". (Note that the residual correlation for Bodily-Kinesthetic tests did not reach the conventional level of statistical significance.) In contrast, the residual correlations between the tests within each of the other three domains–Logical/Mathematical, Spatial, and Linguistic–were noticeably smaller than the corresponding zero-order correlations, thus indicating that much of the shared variance within each of these intelligences was attributable to *g*. Nevertheless, for all three of these domains (especially Linguistic), the residual correlations were large enough to suggest a considerable influence of non-*g* sources of variance on the relations between the tests.

7. Discussion

7.1. Multiple intelligences theory and the structure of abilities

Our analyses of tests measuring the "intelligences" of Gardner's MI theory revealed that many of those tests were substantially intercorrelated, despite representing different domains of Gardner's framework, and also showed strong loadings on a *g* factor and strong correlations with an external test of general intelligence. These results are difficult to reconcile with the core aspects of MI theory. According to that theory, the existence of any such *g* factor would be attributable largely to the verbal demands of the various subtests. However, the tests used in the current study (except those assessing Linguistic intelligence) were selected in such a way as to minimize the demands for high levels of verbal ability. Moreover, both tests representing Gardner's Spatial intelligence–an almost entirely nonverbal domain–showed substantial *g*-loadings.

Therefore, the emergence of a *g* factor defined by purely cognitive tests of several ability domains cannot be attributed to any strong verbal demands associated with those subtests. Instead, it appears that the common element is a general intellectual ability that is expressed especially strongly in reasoning tasks (see Lohman, 2001; Messick, 1992), insofar as the tests with the greatest reasoning demands tended to show the highest *g*-loadings. For example, one of the most highly *g*-loaded tests in this battery was Diagramming Relationships, which clearly demands reasoning ability. In addition, the *g*-loading of Necessary Arithmetic Operations was also very high, and much greater than that of Multiplication and Subtraction, despite the fact that both tests represented Gardner's Logical/Mathematical intelligence domain. The difference in *g*-loadings would appear to be accounted for by the vastly different reasoning demands of these two tests. Tests from domains with relatively little reasoning content, such as Musical and Bodily-Kinesthetic, had much smaller *g*-loadings.

Among the purely cognitive tests, the low correlations involving the Subtraction and Multiplication test are probably attributable to the relatively simple cognitive demands of this task, which required the automatic application of the techniques of arithmetic computation. However, the correlations between the Spatial and Linguistic tests were also weak: even when the two tests within each domain are aggregated, the correlation between these domains reaches only 0.12 (a value that remains unchanged when participant sex is partialled out). This result was somewhat surprising, given that measures of these abilities usually tend to show substantial intercorrelations (e.g., the 0.41 correlation between Block Design and Vocabulary in the normative sample of the Wechsler Adult Intelligence Scale—Third Edition; Wechsler, 2001). The relatively weak correlation observed here is likely attributable in part to the restricted range of ability in the present sample, whose standard deviation in IQ was only 82% of the population level. Nevertheless, the correlations between these domains appear to be atypically low in the present sample, to the extent that the partial correlation between the Spatial and Linguistic domains would actually be *negative* when reasoning abilities are controlled, a result that is not usually observed (Ashton & Vernon, 1995). It is possible that the present sample of participants, most of whom were university students, may include many individuals who are strong either in verbal ability (e.g., humanities students) or in spatial ability (e.g., physical science or visual arts students), or in both of those abilities, but very few individuals who are strong in neither ability. If so, then this would tend to reduce the correlation between the two domains; however, in the absence of data on the levels of verbal and spatial ability of the age-matched, non-university student counterparts of the present participants, it is difficult to test this hypothesis.

The moderate *g*-loadings of the Interpersonal tests suggest that there is indeed a substantial cognitive element to this domain, particularly in light of the fact that the *g*-loading of the Cartoon Predictions test was attenuated considerably by its modest reliability (see Table 3). This finding is consistent with the literature on emotional intelligence, in which it has been reported that this variable, when measured using maximum performance tests, is correlated with tests of general intelligence, but not when measured using self-report inventories (e.g., O'Connor & Little, 2003).

Regarding tests of partly non-cognitive abilities, it is difficult to draw any firm conclusions about the correlations or *g*-loadings involving the two Musical tests, given their extremely poor reliability. The small loadings of the Musical tests and the fact that they were not significantly correlated with the WPT contrasts with some research findings (e.g., Lynn et al., 1989). In addition, whereas studies have found AMMA Tonal and Rhythm to be correlated with spatial tests (Gromko, 2004), no such association appeared in the current study. But again, given the extraordinarily low reliability of the music test used in the present study, the present results do not provide any evidence as to the location of musical ability within the structure of mental abilities. (We suggest that the reliability of the AMMA should be examined in future research, in order to determine whether the present results were specific to this study, or whether they would replicate more broadly.)

The low *g*-loadings of the Intrapersonal tests were as hypothesized, but further exploration of the cognitive content of this domain seems warranted. This study represents a first attempt to assess Gardner's Intrapersonal intelligence through measures based on consistency and accuracy of self-description. Future refinements to the Intrapersonal measures could confirm whether this domain is, indeed, relatively independent of the other ability domains.

The relatively low *g*-loadings of the Bodily-Kinesthetic tests were as hypothesized. With regard to gross motor tasks such as the Stork Stand, there was little reason to expect substantial correlations with tests of any of the other intelligence domains. As for fine motor tasks such as Mark Making, the present study obtained a somewhat lower *g*-loading than has been observed elsewhere (e.g., Hunter, 1983), but our results are consistent with the finding that this ability is less *g*-loaded than are overtly cognitive abilities. Thus, although Gardner has attempted to equalize his eight domains by referring to them all as intelligences, it would seem that Bodily-Kinesthetic ability–or rather, the two or more independent domains of Bodily-Kinesthetic ability–is unlike the other intelligences in that it is relatively lacking in cognitive content.

With regard to the coherence of each of the intelligences, we found that the tests within most ability domains tended to be more strongly correlated with each other than with tests in other domains, suggesting that the intelligences could be viewed as coherent group factors. These findings are unsurprising, and although they are consistent with Gardner's (1983) suggestion of several coherent domains of ability, they are also consistent with the hierarchical models of intelligence that had been developed several decades earlier. Moreover, because the hierarchical models specify an important *g* factor that transcends different types of test content, those models can accommodate the results of the present study much more neatly than does the Multiple Intelligences framework, in which such a *g* factor is not an integral element.

There was some notable variation across domains in the extent to which the two constituent subtests were correlated with each other, independently of *g*. The Linguistic tests, for example, had considerable shared variance beyond their substantial *g*-loadings. The two Naturalistic tests, on the other hand, shared little variance beyond that attributable to *g*, suggesting that Naturalistic intelligence is not a coherent domain. The two Intrapersonal measures were uncorrelated, and the correlation between the two Bodily-Kinesthetic tests was also weak. This finding suggests that these "intelligences" are not well specified and do not constitute coherent ability domains.

7.2. Effects of sample characteristics on observed results

As noted in Results, the present sample had somewhat less variability in general intelligence than is observed in the general population; as a consequence, correlations among the ability tests would tend to be attenuated, thereby leaving a somewhat smaller *g* factor than would be obtained in a sample that is more representative of the population. But in addition to the effects of this restricted range of intelligence, the fact that the sample was above-average in intelligence might also have tended to limit the size of the *g* factor, given the apparent tendency for *g* to account for less variance within high-ability samples than within low-ability samples (e.g., Detterman & Daniel, 1989). Therefore, the characteristics of the present sample are likely to have produced results that underestimate the importance of the *g* factor in summarizing individual differences in performances on the ability tests.

7.3. Practical implications

Without further research, it seems premature for schools and parents to embrace school curriculums based on Multiple Intelligences theory. This recommendation of a cautious approach to the adoption of strategies based on Multiple Intelligences theory does not imply that students should not be treated as individuals with unique ability profiles. However, the findings of the current study do suggest that Multiple Intelligences theory does not provide any new information beyond that already contributed by hierarchical models of ability, and should not be considered a basis for classroom planning.

Future investigations of Gardner's model could focus on the refinement of tests measuring the various domains, particularly Intrapersonal ability tests, and on the use of larger batteries that would allow more complex modeling of the tests' relations with underlying factors. Such investigations might also involve attempts to adopt Gardner's suggestion of assessing individuals by using familiar materials in the performance of culturally-valued tasks, although the development of construct-valid tasks of this kind would likely be rather difficult. In addition, subsequent studies would also profit from the use of participant samples that are more representative of the general population.

8. Conclusions

The substantial *g*-loadings of all purely cognitive tests in the current study contradict Gardner's assertion that there are at least eight independent intelligence domains. Although Gardner has acknowledged the existence of *g* and has conceded that the eight intelligences might not be entirely independent, his contention that positive correlations between various cognitive tasks are largely due to verbal demands was clearly not supported in this study, in which those verbal demands were minimized. Instead, measures of Linguistic, Spatial, Logical-Mathematical, Naturalistic, and Interpersonal intelligences showed a positive manifold of correlations, substantial loadings on a *g* factor, and substantial correlations with an outside measure of general intelligence. The common element that saturated the highly *g*-loaded tests most strongly was their demand on reasoning abilities, not their specifically verbal content.

The finding that several of the partly non-cognitive tests in this study were very weakly *g*-loaded is unsurprising, and suggests that Gardner is likely correct in claiming that Bodily-Kinesthetic ability is quite different from the various cognitive abilities. Given the important contribution of non-cognitive as well as cognitive abilities to performance in the Bodily-Kinesthetic, Musical, and Intrapersonal domains, "talents" might be a more appropriate label than "intelligences".

Finally, some of the ability domains proposed by Gardner were not supported by the present data, as the within-domain correlations were either very weak (Bodily-Kinesthetic, Intrapersonal) or attributable entirely to *g* (Naturalistic). The coherence of some of the other ability domains (e.g., Linguistic, Spatial), as shown by significant residual correlations after the extraction of *g*, is consistent with Multiple Intelligences theory. However, this result is explained equally well by the much older hierarchical models of intelligence, which postulate several group factors in addition to an important *g* factor.

Acknowledgements

This research was supported by the Social Sciences and Humanities Research Council of Canada grant 410-2003-0946. The authors would like to thank Wonderlic Inc. for providing the Wonderlic Personnel Test for our research purposes. The authors are also grateful to Maureen O'Sullivan, who supplied the Social Translations and Cartoon Predictions tests.

References

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, *87*, 707–721.

Ashton, M. C., & Vernon, P. A. (1995). Verbal and spatial abilities are uncorrelated when *g* is controlled. *Personality and Individual Differences*, *19*, 399–401.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, *65*, 546–553.

Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology*, *59*, 538–549.

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

Daz-Lefebvre, R. (2004). Multiple intelligences, learning for understanding, and creative assessment: Some pieces to the puzzle of learning. *Teachers College Record*, *106*, 49–57.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349–359.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests 1976.* Princeton: Educational Testing Service.

Furnham, A., & Fong, G. (2000). Self-estimated and psychometrically measured intelligence: A cross-cultural and sex differences study of British and Singaporean students. *North American Journal of Psychology*, *2*, 191–200.

Gardner, H. (1983). *Frames of mind.* New York: BasicBooks.

Gardner, H. (1993). *Multiple intelligences.* New York: BasicBooks.

Gardner, H. (1999). *Intelligence reframed.* New York: BasicBooks.

Gordon, E. (1989). *Advanced Measures of Music Audiation.* Chicago: GIA Publications.

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Ed.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gromko, J. E. (2004). Predictors of music sight-reading ability in high school wind players. *Journal of Research in Music Education*, *52*, 6–16.

Hunter, J. E. (1983). Overview of validity generalization for the U.S. Employment Service. *USES Test Research Report*, *vol. 43*. Washington, DC: U.S. Department of Labor.

Johnson, B. L., & Nelson, J. K. (1986). *Practical measurements for evaluation in physical education*, (4th ed.) New York: MacMillan.

Lohman, D. F. (2001). Fluid intelligence, inductive reasoning, and working memory: Where the theory of multiple intelligences falls short. In N. Colangelo & S. Assouline (Ed.), *Talent development IV: Proceedings from the 1998 Henry B. and Jocelyn Wallace National Research Symposium on talent development* (pp. 219–228). Scottsdale, AZ: Gifted Psychology Press.

Lynn, R., Wilson, G., & Gault, A. (1989). Simple musical tests as measures of Spearman's *g*. *Personality and Individual Differences*, *10*, 25–28.

McCrae, R. R., Zonderman, A. B., & Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*, 552–566.

Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's Frames of Mind and Sternberg's Beyond IQ in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, *3*, 365–384.

Mettetal, G., Jordan, C., & Harper, S. (1997). Attitudes toward a multiple intelligences curriculum. *Journal of Educational Research*, *91*, 115–122.

O'Connor, R. M., Jr., & Little, I. S. (2003). Revisiting the predictive validity of emotional intelligence: Self-report versus ability-based measures. *Personality and Individual Differences*, *35*, 1893–1902.

O'Sullivan, M., & Guilford, J. P., (1965a). *Cartoon predictions—form A.* Beverly Hills: Sheridan Supply Co.

O'Sullivan, M., & Guilford, J. P., (1965b). *Social translations—form A.* Beverly Hills: Sheridan Supply Co.

Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests. *Journal of Personality*, *66*, 525–554.

Shearer, C. B. (1996). *The MIDAS: A guide to assessment and education for the multiple intelligences.* Columbus, OH: Greyden Press.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences*, *12*, 35–51.

Teele, S. (1992). *Teele inventory of multiple intelligences.* Redlands, CA: Sue Teele and Associates.

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

United States Department of Labor (1970). *General Aptitude Test Battery.* Washington, D.C: US Government Printing Office.

Vernon, P. E. (1961). *The structure of human abilities* (Rev. ed.). London: Methuen.

Wechsler, D. (2001). *Wechsler Adult Intelligence Scale—Third Edition Canadian Technical Manual.* Toronto: Harcourt Canada.

Wonderlic (2002). *Wonderlic Personnel Test and Scholastic Level Exam User's Manual.* Libertyville, IL: Wonderlic.