Intelligence 40 (2012) 560–570

Contents lists available at SciVerse ScienceDirect

Intelligence

Distinguishing verbal, quantitative, and figural facets of fluid intelligence in young students ☆

Joni M. Lakin [a,*], James L. Gambrell [b]

[a] Department of Educational Foundations, Leadership, and Technology, Auburn University, Auburn, AL 36849, United States
[b] Psychological and Quantitative Foundations, The University of Iowa, Iowa City, IA 52242, United States

ARTICLE INFO

Article history:
Received 12 May 2012
Received in revised form 4 July 2012
Accepted 27 July 2012
Available online 24 August 2012

Keywords:
Fluid reasoning
Assessment of children
Bifactor model

ABSTRACT

Measures of broad fluid abilities including verbal, quantitative, and figural reasoning are commonly used in the K-12 school context for a variety of purposes. However, differentiation of these domains is difficult for young children (grades K-2) who lack basic linguistic and mathematical literacy. This study examined the latent factor structure of a picture-based measure of broad fluid reasoning abilities using a bifactor MIRT model to separate general and broad domain factors in a large representative sample of U.S. school children. Substantial evidence showed that picture-based item formats can distinguish between general and domain-specific fluid reasoning abilities in the early school grades. The verbal tasks showed the strongest domain factor and discriminant validity, although the quantitative tasks also showed considerable evidence of a domain factor. Furthermore, comparisons of ELL vs. non-ELL, FRL-eligible vs. non-FRL, Black-White, Hispanic-White, and Asian-White students all yielded small to negligible group differences (below 0.4 *SD*) on these measures. These results compare favorably to differences observed on tests using traditional item formats, and are smaller than the .50–1.0 *SD* group differences often observed in older students.

1. Introduction

Measures of broad fluid reasoning abilities including verbal, quantitative, and figural reasoning are commonly used in K-12 school contexts for a variety of purposes including educational placement, instructional differentiation, and ability-achievement discrepancy assessment (Corno, 1995; Gregory, 2004). Measuring separate verbal and quantitative domains is essential for purposes that require identifying strengths and weaknesses by school subject. The critical challenge in the assessment of these abilities is the measurement of uniquely verbal and quantitative fluid reasoning in young children (grades K-2) who lack basic crystallized linguistic and mathematical literacy skills. The present study evaluated the ability of picture-based item formats to measure distinguishable content factors on a reasoning test. Specifically, we explored the extent to which three broad domain factors were recovered in addition to the overall fluid reasoning factor.

Measuring distinct verbal, quantitative, and nonverbal/figural domains is relatively straightforward for students who are literate in both the language of instruction and mathematical symbol systems (i.e., the majority of students in grade 3 and above). However, verbal and quantitative abilities are more difficult to measure in younger students due to their lack of literacy. For young students, individual administration is a costly solution while many group-administered tests resort to teacher-read item prompts which can create significant demands on students' receptive language skills, thereby introducing construct-irrelevant variance and reducing discriminant validity between domains. Such demands on receptive language (e.g., with English as a common language between student and

☆ The authors gratefully acknowledge the helpful comments of David Lohman on earlier drafts of this article.
* Corresponding author.
E-mail address: joni.lakin@auburn.edu (J.M. Lakin).

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2012.07.005

teacher) make it especially difficult to fairly and validly assess students who are culturally and linguistically diverse (Lohman & Gambrell, 2012).

Although there is some debate on the differentiation of reasoning abilities in young children (Juan-Espinosa, García, Colom, & Abad, 2000; Kane & Brand, 2006; Keith & Reynolds, 2010), researchers have consistently found the presence of strong broad ability factors (beyond *g* or Gf) in all school-aged students (Carroll, 1993; Kane & Brand, 2006). Even in mathematics, which might be expected to depend on schooling and to develop later than other skills, research has shown that quantitative reasoning skills begin to develop before children are taught to count and before they are exposed to formal mathematics education (Starkey, 1992).

1.1. Assessing fluid reasoning across content domains

Fluid reasoning (Gf) can be defined as the process of drawing defensible inferences from incomplete information and it is a central component of cognitive abilities (Carroll, 1993; Gustafsson, 1984; Schneider & McGrew, 2012). Reasoning ability can vary by domain (i.e., the *content* being reasoned about), leading individuals to reason more effectively in some domains than others. The division of fluid reasoning processes by content or domain is supported by extensive empirical research (Beauducel, Brocke, & Liepmann, 2001; Carroll, 1993; Lohman, 2000; Wilhelm, 2005). In Carroll's (1993) compendium of factor analytic studies, three subfactors of fluid reasoning were identified: sequential reasoning, inductive reasoning, and quantitative reasoning. Wilhelm (2005) argues that these reasoning factors may be better understood as content factors rather than process factors with verbal, figural, and numerical content factors defining Gf. This is consistent with Carroll's findings (based on the typical tasks identifying each factor) and is consistent with faceted models of intelligence such as the Berlin Model of Intelligence Structure (BIS) which hypothesize both a content dimension (verbal, numerical, and figural) and a process dimension (reasoning, knowledge, and memory) to intelligence tests (Beauducel et al., 2001; Süß & Beauducel, 2005; Wilhelm, 2005). Beauducel et al. (2001) argue that measuring a content facet in addition to a process facet results in greater construct validity by aggregating the fluid reasoning processes across varying content. In addition to the benefits of aggregation, measuring the three subfactors of Gf also allows the test to align well with the reasoning demands of the typical classroom which includes considerable verbal and quantitative content as well as demands on general abstract reasoning (Corno et al., 2002; Snow & Lohman, 1984; Wilhelm, 2005).

1.2. Purpose of the study

In this study, we explored the construct validity of picture-based measures of verbal and quantitative reasoning for young students (grades K-2). The item formats examined come from the Cognitive Abilities Test Form 7 (CogAT 7; Lohman, 2011). CogAT is a multidimensional (and multi-level) ability test developed for grades K-12 which has a long history of use and well-regarded psychometric properties (DiPerna, 2005; Gregory, 2004). CogAT is one of the most widely used group ability tests in both the United States and the United Kingdom (where a parallel form is used, abbreviated CAT). The test consists of three batteries measuring verbal, quantitative, and figural reasoning with three item formats per battery. In previous editions, the test levels designed for grades K-2 consisted of verbal and quantitative item formats that relied on teacher-read oral prompts with picture-based response options (e.g. CogAT 6; Lohman & Hagen, 2001). This yielded highly correlated verbal and quantitative batteries (Lohman & Hagen, 2002), which was inconsistent with the structure at grades 3–12 (where quantitative and nonverbal/figural correlated more strongly).

CogAT 7 introduced picture-based item formats for young students that were analogous to the verbal and quantitative formats used at higher grade levels. These picture-based formats were designed to draw on conceptual (verbal) reasoning and rudimentary quantitative reasoning. These formats are described in greater detail in the Method section. An added bonus of these formats is the potential to improve the fair and accurate assessment of these abilities for culturally and linguistically diverse students. Because the formats assume little shared language between teacher and student (apart from basic directions), the picture-based items are expected to reduce cultural and linguistic loading. As a result, we expected smaller differences between racial/ethnic, socioeconomic, and language proficiency groups. The authors describe the process used to select "culturally decentered" item content in the Research Guide (Lohman, 2012).

The CogAT 7 picture-based verbal item formats are somewhat similar to formats used by other tests, including the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) and the Kaufman Brief Intelligence Test (K-BIT2; Kaufman & Kaufman, 2004), which both use picture matrix formats. However, the picture-based matrix items on these tests use a mixture of conceptual and visual relationships. Visual relationships (color, size, pattern differences) are most likely to draw on general and figural reasoning rather than verbal reasoning. Only one other test was located that used a pictorial quantitative reasoning format, the Picture Sequence subtest of the Comprehensive Test of Nonverbal Intelligence—Second Edition (CTONI-2; Hammill, Pearson, & Wiederholt, 2009). This test appeared to primarily measure quantitative concepts. However, the test is the only quantitatively oriented task in the battery and contributes to the pictorial rather than a quantitative composite. Thus, in contrast to existing tests, the picture-based formats studied here were designed to require students to identify distinctly conceptual (verbal) and quantitative relationships between the objects represented in the pictures. Through item selection (Lohman, 2012), the tests also diminish the importance of visual features in item solutions.

Though the picture formats are similar to those on the tests above, only CogAT 7 attempts to use them to measure Gf content factors and thus provides the recommended minimum of three indicators for each group factor and three group factors for Gf (Bollen, 1989; Carroll, 1993). Heterogeneity of item content yields a measure of Gf with high "referent generality" (Coan, 1964; Gustafsson, 2002) and less construct underrepresentation (Messick, 1989). This plurality of measures is also helpful for assessing the validity of the new formats in terms of detecting distinct verbal