Intelligence 43 (2014) 52–64

Contents lists available at ScienceDirect

Intelligence

The international cognitive ability resource: Development and initial validation of a public-domain measure

David M. Condon *[,1], William Revelle

Northwestern University, Evanston, IL, United States

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form 11 November 2013
Accepted 7 January 2014
Available online 14 February 2014

Keywords:
Cognitive ability
Intelligence
Online assessment
Psychometric validation
Public-domain measures

ABSTRACT

For all of its versatility and sophistication, the extant toolkit of cognitive ability measures lacks a public-domain method for large-scale, remote data collection. While the lack of copyright protection for such a measure poses a theoretical threat to test validity, the effective magnitude of this threat is unknown and can be offset by the use of modern test-development techniques. To the extent that validity can be maintained, the benefits of a public-domain resource are considerable for researchers, including: cost savings; greater control over test content; and the potential for more nuanced understanding of the correlational structure between constructs. The International Cognitive Ability Resource was developed to evaluate the prospects for such a public-domain measure and the psychometric properties of the first four item types were evaluated based on administrations to both an offline university sample and a large online sample. Concurrent and discriminative validity analyses suggest that the public-domain status of these item types did not compromise their validity despite administration to 97,000 participants. Further development and validation of extant and additional item types are recommended.

1. Introduction

The domain of cognitive ability assessment is now populated with dozens, possibly hundreds, of proprietary measures (Camara, Nathan, & Puente, 2000; Carroll, 1993; Cattell, 1943; Eliot & Smith, 1983; Goldstein & Beers, 2004; Murphy, Geisinger, Carlson, & Spies, 2011). While many of these are no longer maintained or administered, the variety of tests in active use remains quite broad, providing those who want to assess cognitive abilities with a large menu of options. In spite of this diversity, however, assessment challenges persist for researchers attempting to evaluate the structure and correlates of cognitive ability. We argue that it is possible to address these challenges through the use of well-established test development techniques and report on the development and validation of an item pool which demonstrates the utility of a public-domain measure of cognitive ability for basic intelligence research. We conclude by imploring other researchers to contribute to the on-going development, aggregation and maintenance of many more item types as part of a broader, public-domain tool — the International Cognitive Ability Resource ("ICAR").

2. The case for a public domain measure

To be clear, the science of intelligence has historically been well-served by commercial measures. Royalty income streams (or their prospect) have encouraged the development of testing "products" and have funded their ongoing production, distribution and maintenance for decades. These assessments are broadly marketed for use in educational, counseling and industrial contexts and their administration and interpretation are a core service for many applied psychologists. Their proprietary nature is fundamental to the perpetuation of these royalty streams and to the privileged status of trained psychologists. For industrial and

* Corresponding author at: Department of Psychology, Northwestern University, Evanston, IL 60208, United States. Tel.: +1 847 491 4515.
E-mail address: davidcondon2009@u.northwestern.edu (D.M. Condon).

[1] With thanks to Melissa Mitchell.

0160-2896/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2014.01.004

clinical settings, copyright-protected commercial measures offer clear benefits.

However, the needs of primary researchers often differ from those of commercial test users. These differences relate to issues of score interpretation, test content and administrative flexibility. In the case of score interpretation, researchers are considerably less concerned about the nature and quality of interpretative feedback. Unlike test-takers in selection and clinical settings, research participants are typically motivated by monetary rewards, course credit or, perhaps, a casual desire for informal feedback about their performance. This does not imply that researchers are less interested in quality norming data — it is often critical for evaluating the degree to which a sample is representative of a broader population. It simply means that, while many commercial testing companies have attempted to differentiate their products by providing materials for individual score interpretation, these materials have relatively little value for administration in research contexts.

The motivation among commercial testing companies to provide useful interpretative feedback is directly related to test content however, and the nature of test content is of critical importance for intelligence researchers. The typical rationale for cognitive ability assessment in research settings is to evaluate the relationship between constructs and a broad range of other attributes. As such, the variety and depth of a test's content are very meaningful criteria for intelligence researchers — the ones which are somewhat incompatible with the provision of meaningful interpretative feedback for each type of content. In other words, the ideal circumstance for many researchers would include the ability to choose from a variety of broadly-assessed cognitive ability constructs (or perhaps to choose a single measure which includes the assessment of a broad variety of constructs). While this ideal can sometimes be achieved through the administration of multiple commercial measures, this is rarely practical due to issues of cost and/or a lack of administrative flexibility.

The cost of administering commercial tests in research settings varies considerably across measures. While published rates are typically high, many companies allow for the qualified use of their copyright-protected materials at reduced rates or free-of-charge in research settings (e.g., the ETS Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976)). Variability in administration and scoring procedures is similarly high across measures. A small number of extant tests allow for brief, electronic assessment with automated scoring conducted within the framework of proprietary software, though none of these measures allow for customization of test content. The most commonly-used batteries are more arduous to administer, requiring one-to-one administration for over an hour followed by an additional 10 to 20 min for scoring (Camara et al., 2000). All too often, the result of the combination of challenges posed by these constraints is the omission of cognitive ability assessment in psychological research.

Several authors have suggested that the pace of scientific progress is diminished by reliance on proprietary measures (Gambardella & Hall, 2006; Goldberg, 1999; Liao, Armstrong, & Rounds, 2008). While it is difficult to evaluate this claim empirically in the context of intelligence research, the circumstances surrounding development of the International Personality Item Pool ("IPIP") (Goldberg, 1999; Goldberg et al., 2006) provide a useful analogy. Prior to the development of the IPIP, personality researchers were forced to choose between validated but restrictive proprietary measures and a disorganized collection of narrow-bandwidth public-domain scales (these having been developed by researchers who were either unwilling to deal with copyright issues or whose needs were not met by the content of proprietary options). In the decade ending in 2012, at least 500 journal articles and book chapters using IPIP measures were published (Goldberg, 2012).

In fact, most of the arguments set forth in Goldberg's (1999) proposal for public-domain measures are directly applicable here. His primary point was that unrestricted use of public-domain instruments would make it less costly and difficult for researchers to administer scales which are flexible and widely-used. Secondary benefits would include a collaborative medium through which researchers could contribute to test development, refinement, and validation. The research community as a whole would benefit from an improved means of empirically comparing hypotheses across many diverse criteria.

Critics of the IPIP proposal expressed concern that a lack of copyright protection would impair the validity of personality measures (Goldberg et al., 2006). This argument would seem even more germane for tests of cognitive ability given the "maximal performance/typical behavior" distinction between intelligence and personality measures. The widely-shared presumption is that copyright restrictions on proprietary tests maintain validity by enhancing test security. Testing materials are, in theory, only disseminated to authorized users who have purchased licensed access and further dissemination is discouraged by the enforcement of intellectual property laws. Unfortunately, it is difficult to ascertain the extent to which test validity would be compromised in the general population without these safeguards. Concerns about disclosure have been called into question with several prominent standardized tests (Field, 2012). There is also debate about the efficacy of intellectual property laws for protection against the unauthorized distribution of testing materials via the internet (Field, 2012; Kaufmann, 2009; McCaffrey & Lynch, 2009). Further evaluation of the relationship between copyright-protection and test validity seems warranted by these concerns, particularly for research applications where individual outcomes are less consequential.

Fortunately, copyright protection is not a prerequisite for test validity. Modern item-generation techniques (Arendasy, Sommer, Gittler, & Hergovich, 2006; Dennis, Handley, Bradon, Evans, & Newstead, 2002) present an alternate strategy that is less dependent on test security. Automatic item-generation makes use of algorithms which dictate the parameters of new items with predictable difficulty and in many alternate forms. These techniques allow for the creation of item types where the universe of *possible* items is very large. This, in turn, reduces the threat to validity that results from item disclosure. It can even be used to enhance test validity under administration paradigms that expose participants to sample items prior to testing and use alternate forms during assessment as this methodology reduces the effects of differential test familiarity across participants.

While automatic item-generation techniques represent the optimal method for developing public-domain cognitive ability items, this approach is often considerably more complicated