

DATA PAPER

Selected ICAR Data from the SAPA-Project: Development and Initial Validation of a Public-Domain Measure

David M. Condon¹ and William Revelle²

¹ Department of Medical Social Sciences, Northwestern University, Chicago, Illinois, US
david-condon@northwestern.edu

² Department of Psychology, Northwestern University, Evanston, Illinois, US
revelle@northwestern.edu

These data were collected during the initial evaluation of the International Cognitive Ability Resource (ICAR) project. ICAR is an international collaborative effort to develop open-source public-domain tools for cognitive ability assessment, including tools that can be administered in non-proctored environments (e.g., online administration) and those which are based on automatic item generation algorithms. These data provide initial validation of the first four ICAR item types as reported in Condon & Revelle [1]. The 4 item types contain a total of 60 items: 9 Letter and Number Series items, 11 Matrix Reasoning items, 16 Verbal Reasoning items and 24 Three-dimensional Rotation items. Approximately 97,000 individuals were administered random subsets of these 60 items using the Synthetic Aperture Personality Assessment method between August 18, 2010 and May 20, 2013. The data are available in rdata and csv formats and are accompanied by documentation stored as a text file. Re-use potential includes a wide range of structural and item-level analyses.

Keywords: Cognitive ability; intelligence; cognitive ability structure; International Cognitive Ability Resource; ICAR

Funding statement: This research was supported in part by grant SMA-1419324 from the National Science Foundation to William Revelle.

(1) Overview

Collection Date(s)

Data were collected between August 18, 2010 and May 20, 2013.

Background

The SAPA Project is a collaborative online data collection tool for assessing psychological constructs across multiple domains of personality. These domains – temperament, cognitive abilities, and interests – have been chosen based on historical and current prominence in the field of individual differences research. The primary goal of the SAPA Project is to determine the combined and independent structures of each of these domains based on the collection of large, cross-sectional, online samples. Secondary goals include [1] the identification of additional domains (e.g., motivation, character) which may also provide insight into the ways that individuals differ; and [2] an improved understanding of the demographic and psychographic correlates of individual differences in personality.

The data described here were collected to evaluate the item characteristics, reliability and structural properties of ICAR measures assessing 4 distinct components of cognitive ability. The data were also used to validate the ICAR items

based on online administration relative to self-reported achievement test scores and university majors; see Condon & Revelle [1]. The broader goal of these analyses was to demonstrate the utility and potential for public-domain cognitive ability measures that can be administered without proctoring via the internet.

It should be noted that additional ICAR item types have been developed since these data were collected and further development remains ongoing. More information about all of the ICAR measures, including the item content, can be found at the ICAR website (icar-project.com).

(2) Methods

Sample

Participants (N = 96,958) completed the online survey in exchange for feedback about various aspects of their personality and cognitive abilities. No active advertisements or marketing efforts were used to attract participants for this data collection; web traffic statistics (collected through Google Analytics) suggest that participants who did not come to the website directly were directed to it through links from various other websites about personality, personality research, general psychology topics, and psychometrics. Many of these websites were academic/educational in nature.

Many demographic and psychographic variables are included in the data. These include: gender (66.2% of the participants were female); age (see **Figure 1**); marital status (see **Table 1**); body mass index (see **Figure 2**); country (198 countries were represented in total; 78.1% of participants were from the United States, and 34 countries had more than 100 participants); state/region (for 32 countries); educational attainment level (see **Table 2**); employment status (see **Table 3**); self-reported achievement test scores (SAT – Critical Reading, SAT – Mathematics, and ACT Composite); parental educational attainment level (for 1 or 2 parents); and parental field of employment (for 1 or 2 parents). Participants were not required to provide any of these data except age and gender.

Materials

Items from 4 cognitive ability scales were administered. These scales were all developed in the lab of the second author for the purposes of online assessment of cognitive ability. The 4 scales contained a total of 60 items: 9 Letter and Number Series items, 11 Matrix Reasoning items,

16 Verbal Reasoning items and 24 Three-dimensional Rotation items. The Letter and Number Series (“LN”) items prompt participants with short digit or letter sequences and ask them to identify the next position in the sequence from among six choices. Matrix Reasoning (“MR”) items contain stimuli that are similar to those used in Raven’s Progressive Matrices. The stimuli are 3x3 arrays of geometric shapes with one of the nine shapes missing. Participants are instructed to identify which of the six geometric shapes presented as response choices will best complete the stimuli. The Verbal Reasoning (“VR”) items include a variety of logic, vocabulary and general knowledge questions. The Three-dimensional Rotation (“R3D”) items present participants with cube renderings and ask participants to identify which of the response choices is a possible rotation of the target stimuli. None of the items were timed in these administrations as untimed administration was expected to provide more stringent and conservative evaluation of the items’ utility when given online (there are no specific reasons precluding timed administrations of the ICAR items, whether online or offline).

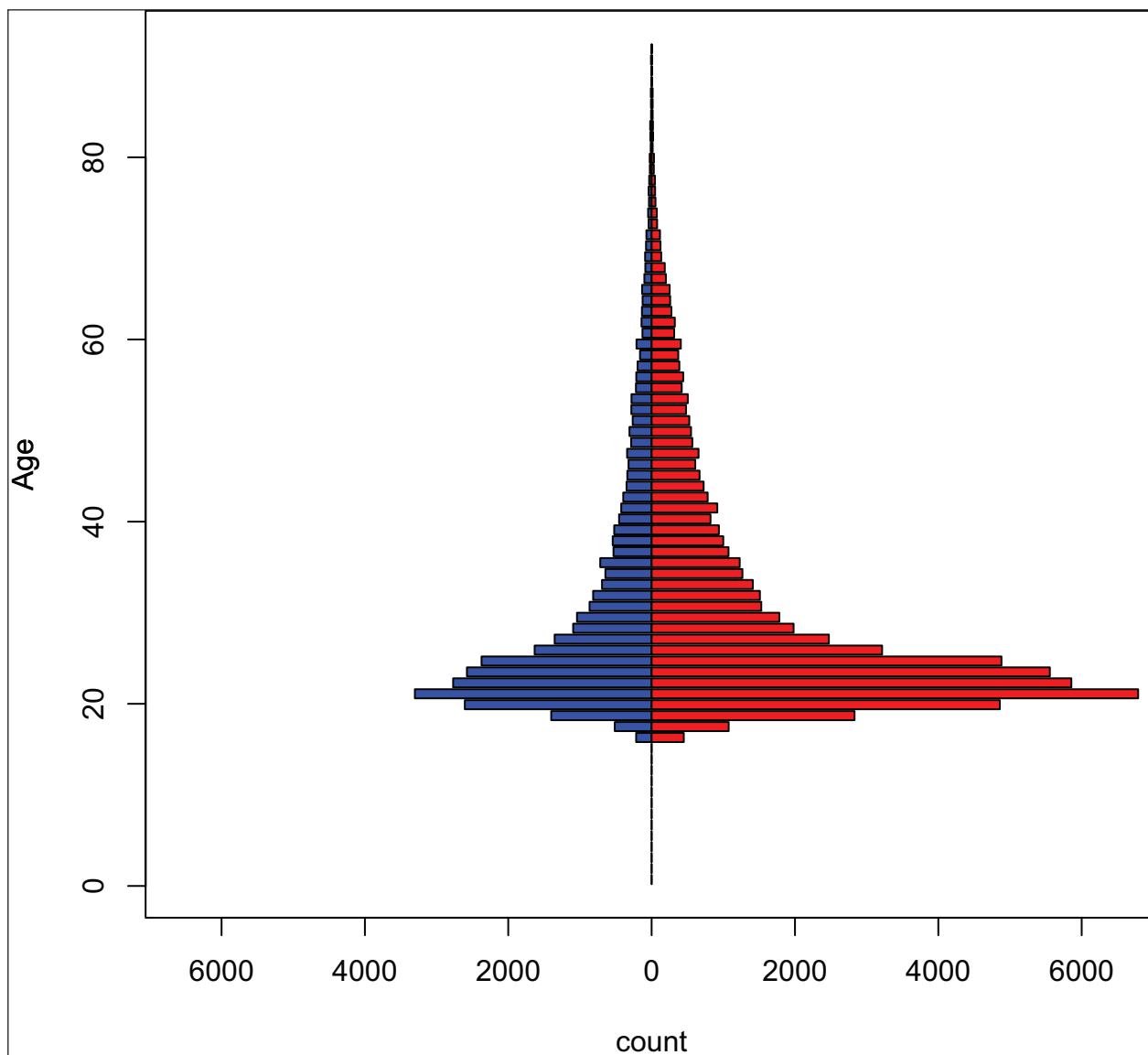


Figure 1: Participants by age and gender (males in blue, females in red).

Marital Status	Participants
Never Married	71,227
Married	16,597
Domestic Partnership	6,070
Divorced & Single	2,470
Divorced & Remarried	497
Widowed & Single	97

Table 1: Marital Status.

For each of the 60 items, participants were instructed to choose the best answer from 8 response choices, including “None of these” and “I don’t know”. In order to maintain the integrity of the measures, the data provided here have already been scored. The raw unscored data may

be obtained by contacting the first author. It should be noted that many useful analyses might be conducted on the unscored data including evaluation of ‘distractor’ response choices and their relationship with both other items and demographic variables.

Procedures

The items were administered using the Synthetic Aperture Personality Assessment (“SAPA”) technique [5], a variant of matrix sampling procedures discussed by Lord [3]. This method produces data which contain “massive missingness” by design [4]. This missingness qualifies for classification as missing completely at random (“MCAR”, 2) and it is further described as massively missing because the mean level of missingness by participant was approximately 68%. The number of administrations for each item varied considerably (median = 21,764; *m* = 19,998;

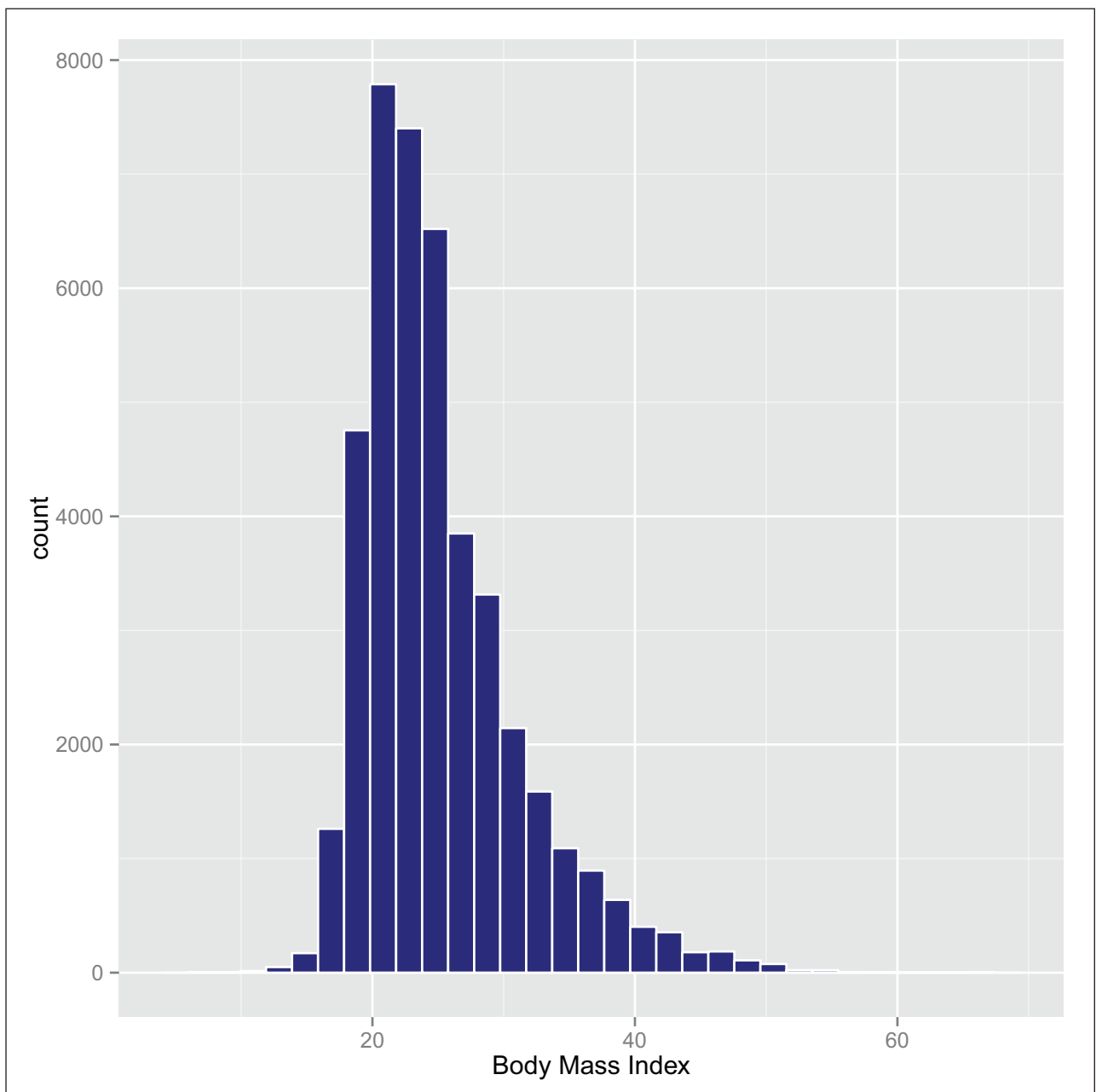


Figure 2: Participants by Body Mass Index.

Educational Level	Participants
Less than 12 years	14,034
High school graduate	5,995
Currently in college/university	49,810
Some college/university, but did not graduate	4,868
College/university degree	11,382
Currently in graduate or professional school	4,225
Graduate or professional school degree	6,644

Table 2: Educational Attainment Level.

Employment Status	Participants
Currently a student	48,716
Employed	37,619
Not employed, seeking work	3,506
Not employed	3,504
Homemaker	1,686
Retired	817

Table 3: Employment Status.

$sd = 10,958$) as did the number of pairwise administrations between any two items in the set (median = 2,610; $m = 4,240$; $sd = 4,110$). The items were presented to participants in random order as part of a broader personality survey, and participants responded to as many items as they wished. The broader survey included items relating to a range of topics such as Big Five personality traits, vocational interests, creativity and more. The number of items administered to each participant was procedurally independent of participant response characteristics; participants were encouraged to complete 16 items. On average, participants responded to 12.4 ($sd = 3.7$; median = 12) of the ICAR items; it is not clear why some participants elected to complete fewer than 16 items though it seems likely that participants skipped the items that were particularly challenging (it would be possible to explore this topic further using the data described here). The feedback provided to participants on these cognitive ability items was informal and basic. Participants were told how many of the cognitive ability items were answered correctly out of those for which a response was given (e.g., “you answered 12 items correctly out of 14”). Participants were also informed about the average number of items answered correctly by previous participants of their age and gender, though no specific interpretative guidance was given about their score. For more information about the development and use of these measures, see Condon & Revelle [1].

Quality Control

The available data are presented largely as they were collected with only two exceptions:

1. Partial removal of data collected from participants who completed the survey more than once in a single browser session. This was done by assigning participants a random user ID that was persistent as long as their current browser session remained active. In those cases where more than 1 response set was entered in a single browser session, only the first response set was kept.
2. Removal of participants with self-reported ages younger than 14 and older than 90. The survey is not intended for participants younger than 14. Self-reported ages over 90 were removed on the grounds that they were deemed to be unlikely.

Ethical issues

No personally identifying information were collected from participants in these data.

(3) Dataset description

Object name

'sapalCARDData18aug2010thru20may2013.rdata'

The data file is named to indicate the data collection method (SAPA), the source of the items (ICAR), and the time period over which the data were collected (18aug2010 through 20may2013). The file can be found at: <http://dx.doi.org/10.7910/DVN/AD9RVY>.

The rdata file includes three objects. The most pertinent of these is the raw data object ('sapalCAR-Data-18aug2010thru20may2013'). The remaining two objects are helper files for data analysis: 'ItemLists' is a list object that provides an index of the ICAR items associated with each item type, and 'superKey60' is a scoring matrix for the 4 ICAR scales (though the items have previously been 'scored' as correct or incorrect - using 1s and 0s, respectively - the scoring key remains useful for scoring the scales). There is also a text file ('demographic codes.txt') which describes the coding for all of the other variables in the data set.

The variable names within the data file have been coded with the acronyms “LN” for Letter and Number Series, “MR” for Matrix Reasoning, “R3D” for Three-Dimensional Rotation, and “VR” for Verbal Reasoning.

Data type

Self-report, cross-sectional survey data from 96,958 participants.

Format names and versions

The data are stored as a single rdata file (approximately 2.7 MB) and three separate csv files (approximately 32 MB in total). The rdata file includes the three objects described above: 'ItemLists', 'superKey60' and the main data object 'sapalCARDData18aug2010thru20may2013'. Each of these three objects are separated into individual csv files. There is also an associated text file that provides full information on the demographic codes ('demographic codes.txt').

Data Collectors

The first and second author were responsible for collecting all the data described in this dataset.

Language

All aspects of the survey and website were written in English. Data collected about the website through Google Analytics suggests that some participants used browser-based translation software, but no specifics are available about the extent and effect of these translations.

License

The data have been deposited under the open license CC0 (Public Domain Dedication).

Embargo

The data are freely available for use with appropriate citation.

Repository location

The data were published on Dataverse and are located at <http://dx.doi.org/10.7910/DVN/AD9RVY>.

Publication date

The dataset was published on September 23, 2015.

(4) Reuse potential

The data are well-suited for many types of structural and correlational analyses of cognitive abilities, including those aimed at reproducing or extending the analysis described by Condon & Revelle [1]. These might include evaluation of the ways in which the 4 item types relate to one another, evaluations of their shared structure, evaluations of structural relationships across constructs in various groups of participants, evaluations of differential item functioning, meta-analyses, or the development of new IRT-based adaptive measures. It should be noted that the feasibility of some of these analyses may be affected by the substantial missingness in the data. Additional, non-overlapping data sets from the SAPA Project are also available for use, including those which contain measures of personality and other constructs; contact the authors for more information.

Competing Interests

The authors declare that they have no competing interests.

Authors Information

The first author is David M. Condon, PhD, an Assistant Professor at Northwestern University's Feinberg School of Medicine in the Department of Medical Social Sciences. His contribution included a primary role in the technical development of the website through which the validation data were collected (sapa-project.org). This role involved the adaptation of existing code (primarily generated by the second author for previous data collection projects) and the authorship of new code (for extending the

functionality and aesthetic design of the website and for improving the data storage and data exportation methods). The first author also took the lead role in cleaning the data to prepare it for sharing, making the data available in the Dataverse, and preparing and submitting this manuscript.

The second author, William Revelle, PhD, is credited with first implementing the survey sampling techniques that made this data collection possible (Synthetic Aperture Personality Assessment, "SAPA") and for developing earlier incarnations of the survey described on his website (personality-project.org). He also owns the url for the website where these data were collected (sapa-project.org). The second author played a secondary role in the development and maintenance of the website used to collect these data.

The first and second author both contributed substantially to the design, development and psychometric validation of the ICAR item types included here. Melissa Mitchell (not listed as an author) also contributed to the original design of several items.

Acknowledgements

The authors would like to acknowledge Melissa Mitchell for her contribution to the second author in developing several of the ICAR items.

References

1. **Condon, D M** and **Revelle, W** 2014 The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, 43: 52–64. DOI: <http://dx.doi.org/10.1016/j.intell.2014.01.004>
2. **Graham, J W** 2009 Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, 60(1): 549–576. DOI: <http://dx.doi.org/10.1146/annurev.psych.58.110405.085530>
3. **Lord, F M** 1955 Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 20(1), 1–22, DOI: <http://dx.doi.org/10.1007/bf02288956>
4. **Brown, A D** 2015 (July) Standard Errors of SAPA Correlations: A Monte Carlo Analysis. In W. Revelle (Chair), *Studying Individual Differences Using the Web: A Report from the SAPA Project*. Symposium conducted at the biennial meeting of the International Society for the Study of Individual Differences, London, Ontario, CA.
5. **Revelle, W, Wilt, J** and **Rosenthal, A** 2010 Individual Differences in Cognition: New Methods for Examining the Personality-Cognition Link. In Gruszka, A, Matthews, G and Szymura, B (Eds.) *Handbook of Individual Differences in Cognition*. New York, NY: Springer New York, pp. 27–49. DOI: http://dx.doi.org/10.1007/978-1-4419-1210-7_2

How to cite this article: Condon, D M and Revelle, W 2016 Selected ICAR Data from the SAPA-Project: Development and Initial Validation of a Public-Domain Measure. *Journal of Open Psychology Data*, 4: e1, DOI: <http://dx.doi.org/10.5334/jopd.25>

Published: 26 January 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Journal of Open Psychology Data* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS 