BUROS CENTER FOR TESTING

TEST REVIEWS

Assessment of Multiple Intelligences



Assessment of Multiple Intelligences

Purpose

Designed to "assess which of Gardner's Eight Intelligence Types a person possesses."

Population

Below age 17 to over age 40.

Publication Date

2011.

Acronym

AMI.

Scores, 8

Bodily-Kinesthetic, Logical-Mathematical, Linguistic, Visual-Spatial, Musical, Intrapersonal, Interpersonal, Naturalistic.

Administration

Group.

Price Data

Available from publisher.

Time

(20) minutes.

Comments

Self-administered online assessment. The test publisher advises that the test manual is being updated to include more information about methodology and theoretical background used in the development of the test. The test publisher also advises that this information is available to clients, as are benchmarks for relevant industries and racial/ethnic group comparison data. However, this information was not provided to Buros or the reviewers.

Author

PsychTests AIM, Inc.

Publisher

PsychTests AIM, Inc. [Canada].



REVIEW 1 OF 2

Review of the Assessment of Multiple Intelligences by GARY L. CANIVEZ, Professor of Psychology, Department of Psychology, Eastern Illinois University, Charleston, IL:

DESCRIPTION

The Assessment of Multiple Intelligences (AMI) purports to measure "which of Gardner's eight intelligence types a person possesses" for use in pre-employment testing (manual, p. 1). The AMI is also "designed to facilitate improvement and increase the test-taker's suitability for this career" (p. 1). Unfortunately there is no description as to how the AMI is to be used in pre-employment testing or otherwise. It is computer administered and scored based on a self-report questionnaire or survey with 55 questions and presents results and scores for each of Gardner's eight hypothesized intelligences: Bodily-Kinesthetic, Logical-Mathematical, Visual-Spatial, Linguistic, Musical, Intrapersonal, Interpersonal, and Naturalistic. The AMI contains nine questions asking the examinee to indicate which response (out of four to nine options) best answers a question about what they would or might do, think, or remember, or about how others might perceive them; six questions that ask the examinee to check all activities, jobs, hobbies, previous best school subjects, venting feelings, and rating strengths that apply; one question with a list of careers they think they could be best at with adequate resources and training from which they select three; and 39 statements or characteristics rated on a 5-point scale ranging from exactly like me to not at all like me. Self-report responses related to each of the eight intelligences produce scores, but there is no indication of the type of score reported (raw score, standard score, percentile rank, etc.). It appears that the results are simply raw scores totaling item responses related to the particular intelligence dimensions.

A computer-generated report is produced that first summarizes the individual's "main" intelligence type, specified by the intelligence scale that results in the highest score. There is no explication in the manual as to how this "main" type is determined when there is not a significant difference between the highest and the next highest score; presumably the "main" intelligence type is simply the one with the highest score. The report includes a description of Gardner's theory and beliefs, followed by a graph, reflecting a profile, of the individual's scores on the eight intelligence scales. No standard error of measurement-based confidence intervals are presented on the graph or elsewhere. A presentation of each of Gardner's eight intelligences with a description, motto, list of common abilities/strengths/interests, and list of famous people who presumably represent that intelligence is included with the individual's scores),



"Influencing Intelligence Types" (those with "medium" scores), and "Least Developed Intelligence Types" (those with "low" scores). However, like the determination of the "main" intelligence type, the manual provides no information or criteria for what constitutes a high, medium, or low score. Finally, the report ends with "Advice" regarding each of the eight intelligences presented with "Tips" on how to maximize that type (presumably by practicing or doing more of it). The vast majority of the report appears to be canned statements that do not differ regardless of responses. There are no peer-reviewed research publications or books referenced in the report or the manual.

DEVELOPMENT

The manual provided for this review provides little or no information about how the AMI was developed or about the procedures and iterations of item construction or evaluation. No rationale is provided for why a self-report survey was used in place of directly assessing intellectual skills and abilities through actual performance. There is no report of expert evaluation of essential content to guide construct representation, and there is no report of pilot testing, item analyses, item refinement, or criteria for decisions to add, delete, or modify items or item types. Statistical analyses and tables are included in the manual, but there is no description of a priori hypotheses stated for statistical comparisons across gender, age, education, grades, or occupational position that would guide interpretation of statistical results. The test manual presents statistically significant findings, but no discussion of the meaning or implications of these differences is provided.

TECHNICAL

Statistical information is presented in the manual, but there is no normative sample or transformation of raw scores to standardized scores. A total of 10,135 individuals (5,463 [54%] female, 3,595 [35%] male, 1,077 [11%] unknown gender) reportedly completed the AMI on Queendom.com, Psychtests.com, or PsychologyToday.com. The sample was uncontrolled, as respondents self-selected to participate in the validation data collection. Reported frequencies within age categories indicated 29% were below 17, 29% were 18-24, 11% were 25-29, 12% were 30-39, 11% were 40 and above, and 8% did not report their age. Disappointingly, descriptive statistics for age are not provided, so it is unknown how young the youngest or how old the oldest participants were. Also there is no indication why the specified unequal age ranges were selected or whether individuals reported exact ages that were later converted to age ranges.

Descriptive statistics for the eight intelligences include minimums, maximums, means, and standard deviations, but skewness and kurtosis indexes are not reported. Histograms presented for each of the eight intelligence dimensions illustrate that all eight were negatively skewed and indicate that participants tended to rate themselves more favorably on all dimensions. This is in contrast to performance-based intelligence tests in which subtests and composite scores are normally distributed. The manual contains

All Mental Measurements Yearbook test reviews are copyrighted by the Buros Center for Testing. Reviews may be printed for individual use only, and may not be otherwise duplicated or distributed without written permission. Please refer to the Buros website for further information about copyright and permissions: <u>http://www.buros.org/copyright-and-permissions</u>. **4**



tables of descriptive statistics of scores by gender, age, education, grades, and occupational position. Missing is any description of participant race, ethnicity, geographic region, community size, or country of origin.

The only method of reliability reported for the eight intelligence scale scores is internal consistency. Alpha coefficients for the scale scores ranged from .64 (Naturalistic) to .79 (Logical-Mathematical). The alpha coefficient for the overall scale was .82, even though there is no total or overall score provided by the AMI. The sample size and sample characteristics for these calculations were not disclosed. These internal consistency estimates are substantially lower than the criterion (.90) many measurement experts indicate is necessary for individual test use or clinical decision-making (Nunnally & Bernstein, 1994; Ponterotto & Ruckdeschel, 2007; Salvia & Ysseldyke, 2001). Although alpha coefficients are presented, no standard errors of measurement or confidence intervals (obtained score or estimated true score) are provided in either the manual or the report. Standard errors of measurement are necessary for proper test score reporting and interpretation (AERA, APA, & NCME, 1999). Because intelligence is a construct that is assumed to be stable across time, it is further disappointing that short-term test-retest (stability) estimates are not provided as is customary.

Prior to reporting what were described as validity analyses, t-test and ANOVA procedures were noted, and sample size was acknowledged to be an issue in interpreting statistically significant differences that might not have practical importance. Presentation of effect size estimates (i.e., Cohen's d; Cohen, 1988) would have easily illustrated such differences, but the manual is devoid of these important estimates. It was further noted that when the number of subjects in validation samples were not equally balanced, a smaller random sample from the larger group was made to "conduct the analyses effectively" (p. 3). Reductions in sample sizes were apparently made for females, individuals under age 25, and in an unknown manner across education, grades, and occupation. Resulting sample sizes across subgroups were not identical but were closer than those in the total sample. There are no multivariate analyses (MANOVA) or factorial analyses examining interactions between the five validity variables on the eight intelligence dimensions.

It was noted in the manual that all data provided were self-reported; there was no report of an attempt or ability to verify the accuracy of data obtained. Participants were asked to reply to questions regarding their gender, age, educational level (10 categories from grade school through Ph.D./Doctoral Degree), grades (top, good, average, or below average), and position (entrepreneur, managerial, non-managerial, or not employed). The manual notes that age, education, grades, and position variables were "recoded" but this adjustment is not explained. Presumably these variables were used to test the validity of AMI scores, but there is no discussion about expected differences or hypotheses for differences between various groups. Only t-tests (gender) or univariate ANOVAs (age, education, grades, position) and descriptive statistics are presented; given that the AMI includes eight scales (eight dependent variables), multivariate analyses should have been reported first. Significant (p < .05) or "marginally" significant (p < .10) post hoc

All Mental Measurements Yearbook test reviews are copyrighted by the Buros Center for Testing. Reviews may be printed for individual use only, and may not be otherwise duplicated or distributed without written permission. Please refer to the Buros website for further information about copyright and permissions: <u>http://www.buros.org/copyright-and-permissions</u>. **5**

BUROS CENTER FOR TESTING

comparisons are reported, but effect sizes are not presented. Following presentation of statistically significant differences across subgroups, bar graphs of group means are presented to illustrate the subgroup differences. To judge these results, this reviewer calculated effect sizes based on the largest mean differences reported in the tables in the manual across the five comparision variables. The majority of comparisons reported to be statistically significant were of trivial or small effect sizes (Cohen, 1988) and thus not of practical or clinical importance. Several medium effect sizes were found, but no mean comparisons achieved a large effect size. Particularly troubling is the fact that mean differences between those reporting Ph.D./Doctoral Degrees and those reporting Some High School produced Cohen's d effect sizes of only .59 (Logical-Mathematical), .68 (Linguistic), and .032 (Visual-Spatial); other effect sizes were much smaller, including for those reporting only a Grade School education. It is also clear that the graphs in the manual illustrating group differences are presented in a manner that grossly exaggerates group differences visually because of stretching out the vertical axis and starting it well above zero.

COMMENTARY

Tests and assessment instruments are constructed in a manner that reflects both theoretical constructs and practical applications. Instruments have specific applications, and psychometric research is conducted to provide evidence for the various uses and interpretations proposed by authors and publishers. In addition to a test demonstrating strong evidence for reliability, validity, and utility of scores, there must be a detailed manual that provides ample evidence for the theoretical background and research supporting the construct being measured; explicit description of the methods and procedures of test construction including item development, analysis, and modification; detailed description of a normative sample from which norm referenced scores would be derived based on a representative sample from the population; detailed description of various methods of psychometric evaluation of reliability, validity, and utility; and, finally, detailed description of interpretation procedures for application of scores for groups and individuals. The AMI and its manual fall far short of approximating any of these important qualities.

With respect to the underlying theory of the AMI, there are many critics of Gardner's theory, and many have pointed out major shortcomings (e.g., Barnett, Ceci, & Williams, 2006; Brody, 2006; Jensen, 1998; White, 2006). Generations of psychologists have studied intelligence for more than a century and have measured and quantified aspects such as quantitative reasoning (Logical-Mathematical), verbal reasoning (Linguistic), and spatial visualization (Visual-Spatial). These aspects seem to be the only intelligences in Gardner's theory with strong evidence and that others would consider intelligence. Although other abilities or characteristics Gardner calls intelligence do constitute differences across individuals, very few agree with Gardner that they are intelligence or should be considered so.

Messick (1989, 1995) argued for validity of tests to be examined along a number of important dimensions and to be based on evidence from (a) test content, (b) response processes, (c) internal structure, (d)

All Mental Measurements Yearbook test reviews are copyrighted by the Buros Center for Testing. Reviews may be printed for individual use only, and may not be otherwise duplicated or distributed without written permission. Please refer to the Buros website for further information about copyright and permissions: <u>http://www.buros.org/copyright-and-permissions</u>.



relations with other variables, and (e) consequences of testing. These areas are specifically noted in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999). The only reported approach to assessing validity of AMI scores was with relations with other variables related to general comparisons across demographic subgroups without specified hypotheses or expectations from theory or past research. There were no correlations between the eight AMI scales presented, no internal structural analyses (exploratory or confirmatory factor analyses), no predictive validity analyses, and no diagnostic efficiency or utility analyses. There were no comparisons of the AMI self-ratings with objective measures of the actual skills of the individuals. Without such data it is impossible to meaningfully interpret AMI scores. There are no inferences one can make based on AMI scores without such validity data. The available evidence in the AMI manual showing inadequate reliability and the general dearth of information in the manual is not consistent with Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), and the AMI should not be used in individual clinical decision-making. In the words of Weiner (1989), the ethical psychologist will "(a) know what their tests can do and (b) act accordingly" (p. 829). In the case of the AMI, this test (a) has not demonstrated acceptable levels of reliability, validity, or diagnostic utility of scores and (b) ought not be used in individual assessment situations, pre-employment or otherwise.

SUMMARY

The AMI was reportedly designed to assess which of Gardner's eight intelligence types individuals possess, to be used in pre-employment testing, to suggest improvement, and to increase the individual's career suitability. Based on the extreme lack of information provided in the AMI manual and the apparent lack of published peer-reviewed research, it appears that the AMI can do none of these. It is strongly recommended that the guidance of Clark and Watson (1995) and Standards for Educational and Psychological Testing (1999) be followed and that the AMI not be used in any clinical, pre-employment, or other individual evaluation situations until it is revised and refined and strong reliability, validity, and utility evidence is provided. Individuals who complete the AMI and the professionals attempting to make sense of and use scores from the AMI are sadly provided scores that are not normatively based, are insufficiently reliable, disregard error in measurement, lack validity data to make interpretations, and lack utility data regarding whether scores assist in correctly identifying or classifying individuals. At best, the AMI is a novelty test like those commonly found on Internet sites. Should one be interested in assessing intelligence then one of the many well normed, standardized, and research abundant skill-based measures of intellectual or cognitive abilities should be used.

REVIEWER'S REFERENCES

American Educational Research Association, American Psychological Association, & National Council on



Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Barnett, S. M., Ceci, S. J., & Williams, W. M. (2006). Is the ability to make a bacon sandwich a mark of intelligence?, and other issues: Some reflections on Gardner's theory of multiple intelligences. In J. A. Schaler (Ed.), Howard Gardner under fire: The rebel psychologist faces his critics (pp. 95–114). Chicago, IL: Open Court.

Brody, N. (2006). Geocentric theory: A valid alternative to Gardner's theory of intelligence. In J. A. Schaler (Ed.), Howard Gardner under fire: The rebel psychologist faces his critics (pp. 73–94). Chicago, IL: Open Court.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7, 309–319.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 741–749.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York, NY: McGraw-Hill.

Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. Perceptual and Motor Skills, 105, 997–1014.

Salvia, J., & Ysseldyke, J. E. (2001). Assessment (8th ed.). Boston, MA: Houghton Mifflin.

Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. Journal of Personality Assessment, 53, 827–831.

White, J. (2006). Multiple invalidities. In J. A. Schaler (Ed.), Howard Gardner under fire: The rebel psychologist faces his critics (pp. 45–71). Chicago, IL: Open Court.



Cite this review

Canivez, G. L. (in press). [Test review of Assessment of Multiple Intelligences]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook*. Retrieved from http://marketplace.unl.edu/buros/

REVIEW 2 OF 2

Review of the Assessment of Multiple Intelligences by ELEANOR E. SANFORD-MOORE, Senior Vice-president of Research and Development, MetaMetrics, Inc., Durham, NC:

DESCRIPTION

The Assessment of Multiple Intelligences (AMI) was developed and published by PsychTests AIM, Inc. in 2011. The test is based on the work of Howard Gardner and is designed to assess "the manner in which a person learns best. It will identify which specific type of intelligence an individual possesses and how it can be used to his or her advantage" (ArchProfile catalog, 2012). The test reports scores related to eight intelligence types: Bodily-Kinesthetic, Logical-Mathematical, Visual-Spatial, Linguistic, Musical, Intrapersonal, Interpersonal, and Naturalistic. Scoring reports also provide advice "to facilitate improvement and increase the test-taker's suitability for this career" (manual, p. 1). In addition to the eight scores reported, the scores are categorized as Dominant Intelligence Types, Influencing Intelligence Types, and Least Developed Intelligence Types.

The AMI consists of 55 items and is administered online. The items include selected-response (Likert-type) items in which the examinee is asked to indicate the extent to which the statement describes him or her (self-assessment) and items in which the examinee is asked to select all of the statements that describe what he or she would do in a specific situation (situational). The test takes about 20 minutes to complete. The AMI is scored by computer and a printable report is provided to the test-taker with detailed information related to the descriptions of each of the eight scores and associated common capacities and strengths. No information concerning the scoring procedures is provided in the test manual.

DEVELOPMENT

The Assessment of Multiple Intelligences is based on the work of Howard Gardner and his Multiple Intelligence Theory. This theory was described in 1983 as a classical model that could be used to understand and teach many aspects of human intelligence, learning styles, personality, and behavior. The

BUROS ENTER FOR TESTING

model has been used in both education and industry to provide an indication of people's preferred learning styles, as well as their behavioral and work styles, and their natural strengths.

No information is provided in the technical manual about the development of the items or the assessment. One table provides the number of items per scale, but no information is provided showing how the 55 items are incorporated into each of the eight scores (intelligence types). It is obvious from the table that most, if not all, of the items contribute to numerous scores (the number of items per scale ranges from 22 to 23). No information is provided as to the factor structure of the eight scores. Also, no information is provided related to the scaling of the eight scores and the criteria for the three categories of the scores.

TECHNICAL

No information is provided related to standardization of the results or norming populations.

Reliability was calculated using coefficient alpha. The reliability estimate of the overall scale was .82, and the reliability estimates of the eight scores (intelligence types) ranged from .64 to .79 across one study (N = 10, 135; examinees self-selected to take the assessment and opted in to participate in the study). No information related to test-retest reliability is presented.

The validity information presented in the manual consists of one study of 10,135 self-selected examinees. The distributions of each of the eight scores are negatively skewed with means between 70 and 80 (range of scores is from 0 to 100). The scores for the eight intelligence types are compared across various subgroups: gender, age, education, grades, and [employment] position (subgroup information provided by participants). No information was provided as to where differences should be expected. For the variable of gender, all differences between men and women were significant. For the characteristic age, a variety of comparisons for all eight scores were significant. Across levels of education, a variety of significant comparisons were observed for four of the eight scores (Logical-Mathematical, Linguistic, Intrapersonal, and Interpersonal). Comparing grades, a variety of significant comparisons were observed for seven of the eight scores (the exception being Musical). A variety of significant comparisons were observed for seven of the eight scores (the exception was Visual-Spatial) when vocational positions were compared.

COMMENTARY

The results from only one study were presented in the technical manual, and no interpretive information about the results was provided. In order for the test results to be useful, information should be provided within the developmental model for the assessment to explain when significant comparisons should be expected. An Internet search finds little empirical evidence to support the eight intelligence types in the

All Mental Measurements Yearbook test reviews are copyrighted by the Buros Center for Testing. Reviews may be printed for individual use 10 only, and may not be otherwise duplicated or distributed without written permission. Please refer to the Buros website for further information about copyright and permissions: http://www.buros.org/copyright-and-permissions.



Gardner model. Given the high degree (inferred) of overlap of items across the eight scores on the AMI (55 items total and each of the eight scores is based on 22 to 23 items), there does not seem to be support for all eight scores. In addition, although numerous subgroup comparisons were statistically significant, the actual observed score differences between the highest and lowest subgroups were often small (less than one-fourth of a standard deviation).

SUMMARY

The Assessment of Multiple Intelligences is administered and scored online. The technical manual provides limited psychometric information for the AMI and provides no information related to the development of the items and the assessment, scoring, or construct validation. Although the results do have "face validity" when comparing the scores and reading through the descriptions of the scores, construct validity information should be provided to support eight scores and the use of the results to "provide insight about what is important to the examinee and/or which settings the examinee would most likely thrive" (from online directions for AMI).

Cite this review

Sanford-Moore, E. E. (in press). [Test review of Assessment of Multiple Intelligences]. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook*. Retrieved from http://marketplace.unl.edu/buros/