Analysis of the Validity and Reliability of the Das-Naglieri CAS (Cognitive Assessment System)-Arabic Edition

Nazeh Natur

Al-Qasemi Academic College, Baqa, Israel

The present study was carried out in Israel in order to examine the validity of the translated CAS (Cognitive Assessment System) Basic Battery for Arabic speaking school aged children in Israel, ages 5.0-7.11 years on a sample population of 49 (21 males, 28 females) students. Data collected were cross-validated with data obtained from a matched sample of English-speaking, American students (retrieved from the test standardization sample, using English version). Analysis of the internal consistency of the CAS-Arabic Edition indicated a high degree of reliability; comparison between performance of the groups scores and revealed a high degree of validity. Further analysis of data on the impact of cross-language translation, age, gender, and parental education attainment confirmed that the CAS-Arabic battery is psychometrically appropriate for use with this distinct ethnic group of children in Israel.

Keywords: CAS (cognitive assessment system) Arabic, Arabic assessment, CAS (cognitive assessment system), psychologocal assessment

Introduction

Despite their widespread clinical and educational value in Western societies, cognitive assessment tools remain under-developed, under researched, and underutilized in many non-Western societies, particularly in Arabic-speaking societies (Ahmed & Geilen, 1998). Indeed, while a number of traditional cognitive batteries have been translated and normed, such as the WISC-IV-Spanish, K-ABC-1984, for use with Spanish-speaking populations both within and outside of the United States, no contemporary tests of cognitive ability have been developed for use with Arab-speaking populations (Ahmed & Geilen, 1998). Outside of Israel, an examination of extant literature in the Arab world reveals that research on standardized assessments of cognitive abilities is almost nonexistent in Middle-Eastern, Arabic-speaking societies. Most work on the development of standardized psychological tests in Arab society is outdated and has relied heavily on translated tests from Western societies, most notably the United States (Ahmed & Geilen, 1998). In 1934, two Arab psychologists, El- Kappani and El-Koussy, were credited with the pioneering work of translating, norming, and standardizing Western-based psychological tests for use with Arab populations in the mid-1930s and 1940s. Other Arab scholars, most notably L. K. Meleika, pursued this line of work, including the adaptation and standardization of the original Stanford Binet Intelligence Test (Ahmed & Geilen, 1998), the Wechsler Adult Intelligence Scale, original version (Ahmed & Geilen, 1998), and the MMPI (Minnesota Multiphasic Personality Inventory),

Nazeh Natur, Ph.D., associate professor, Faculty of Education, Department of Special Education and Psychology, Al-Qasemi Academic College Baqa.

original version (Khaleefa & Ashria, l995). However, many of the existing Arabic-translated Western tests are out of date, are not in line with current testing standards, use outdated or inappropriate norms, and are not readily available for practice by psychologists and special educators (Khaleefa & Ashria, l995). Furthermore, several Middle-Eastern scholars have recently expressed the urgency and significant need for increased research in the development, validation, and standardization of Arabic-based cognitive assessment measures (Amer, 2007).

The status of cognitive assessment in Arab-Israeli society, whose population is the focus of the current study, mirrors that of the larger Arab world. Although Israel is considered by many to be one of the more Westernized societies in the Middle East, the development and utilization of standardized, Arabic language measures of cognitive ability in Israel has been extremely limited and research in this area continues to be extremely sparse (Zeinder, 2004). In contrast, the state of cognitive testing in Jewish Israel is comparable in its use and functions to that of the United States where it is used for selection purposes, classification, and placement of students (Beller, 1992). In addition to the cognitive intellectual tests that have been translated and adapted to the Israeli Jewish population (e.g., WPPS (The Wechsler Preschool and Primary Intelligence), Lieblich, 1975; WISC-R (The Wechsler Intelligence Scale for Children-Revised), Cahan, 1998; and K-ABC (The Kaufman Assessment Battery for Children), Cahan & Noyman, 2001), there is the groundbreaking work of Feuerstein and the development of the Dynamic Assessment System, which provides assessment and systematic interventions for modifying the cognitive competencies of special education children (Feuerstein, Randy, Hoffman, & Miller, 1980). While these cognitive tests have been adapted and standardized on the Israeli-Jewish, Hebrew-speaking population, such measures are nonexistent for the Israeli-Arab population, which constitutes 20% of Israel's population (Coursen-Neff, 2005).

With emphasis placed on education in the Arab world and with the intensity of modernization and globalization, there is an urgent need to develop assessment tools that would allow appropriate education to all members of the society. Lack of proper, standardized assessment tools made available to the Arab population means misdiagnosis of special needs children, placement in inappropriate settings, and finally, graduating a large portion of school-age children without basic education. This study is the first effort in this direction and focuses on validating and adapting this instrument to populations most in need of such measures: the Arabs in Israel. Furthermore, validation of this instrument would be the first step towards standardization and possible future expansion in the field of psychometrics in Arabic. The great utility of standardized, norm-referenced tests is its ability to make a valid comparison of performance of an individual to his or her peer group. They provide some degree of quantification of the child's psychological functioning in the present, in reference to a peer group. In essence, they provide a snapshot of the child's weaknesses and strengths in motor, behavioral, and cognitive functioning, and provide a baseline that assists in measuring the progress of the child in response to a planned intervention. Norm-referenced tests also allows for the evaluation of physical, social, neurological, cognitive, and developmental changes in a child's condition through the standardization of instructions, testing conditions, and relatively objective scoring guidelines (Sattler & Hoge, 2004).

On the educational level, the purpose of assessment includes screening, problem-solving, diagnosis, counseling and rehabilitation, and progress evaluation (Sattler & Hogue, 2004). American Federal laws as well as Israeli Law for special education require screening and identification of children according to their abilities in order to be appropriately placed in the educational system.

A screening assessment is a brief assessment that aims to identify children's eligibility for certain

programs, or to provide remediation to children who have been identified with some disorder. A problem-solving assessment is a more detailed evaluation of a specific area of functioning. It is usually done to address a diagnostic or skill question. A diagnostic and classification assessment, on the other hand, is a detailed evaluation of the individual's strengths and weaknesses in the areas of cognitive, academic, language, and social functioning. In this assessment, recommendations on interventions and placement are offered. Similar to the diagnostic assessment is the rehabilitation assessment, except that it focuses on the child's ability to adjust and to succeed on daily basis. A progress evaluation assessment focuses on the periodical progress of the child and evaluates developmental changes and appropriateness of interventions in place. On the societal level, Western culture is achievement driven. People with higher intelligence seem to be more successful in society (Hunt, 1995). Complex jobs require a higher level of mental tasks and performance than less complex jobs. Thus, cognitive assessments are used on many levels in business and job placements (Hunt, 1995).

The importance of conducting this study originated from the dire need to identify culturally sensitive and technically adequate nonverbal measures of ability for Arabic language speakers, and to adapt and validate a cognitive test which could be used with the Arab school-aged population. The present study examines the utility of the Arabic Das-Naglieri CAS (Cognitive Assessment System)-Arabic Edition by evaluating its validity. Specifically, the study attempts to achieve tow objectives: (1) to examine the psychometric properties of CAS among a sample of 49 Arabic-speaking Palestinian students in Israel; and (2) to cross-validate it with a matched sample of 49 English-speaking students in the United States, retrieved from the standardization sample.

The CAS represents a major cognitive assessment battery that has been widely used in Western societies (Bracken & McCallum, 1999). CAS focuses on assessing an individual's intellectual functioning and providing information about the individual's cognitive strengths and weaknesses, while associating the weaknesses or strengths to one of four areas: planning, attention, simulation, and successive. It allows flexible administration and can be offered in two different batteries: the standard battery with 12 subtests or the basic battery with eight subtests. Either one is suitable for ages of 5.0-7.11 and 8.0-17.11. Finally, like other validation studies on cross-culture, this efforts enhances the cross-cultural validation for the underlying construct of the CAS theory psychological processes which ultimately increase our understanding of the salient and/or universal characteristics of particular cognitive tests batteries, which, by extension, might also lend further support for (or refutation of) the validity of the particular cognitive theory on which the test is based. Cross-cultural validity studies can lead to significant improvements in the quality of care and educational planning and opportunities available to previously untested members of a society. CAS is minimally verbal and culturally sound for its use of geometric matrices. CAS-Arabic Edition required minimum changes resulted from administering a pilot study. Still, a rigorous translation and adaptation guidelines were followed as recommended (Hambleton, 1993; Herdman, Fox-Rushby, & Badia, 1997).

In essence, this study advocates for the new development of culturally fair intellectual measures that are psychological process driven rather than general ability driven to better assess intellectual abilities. Validated in Arabic, the CAS-Arabic Edition will serve as a reliable measure and a resource for mental health professionals serving this community, policy makers, and other key stakeholders.

In the current study, it was postulated the performance of Arabic-speaking students on CAS-Arabic Edition will be comparable to the performance of both groups: (1) matched sample of English-speaking, American students (retrieved from the test standardization sample); and (2) the original CAS standardization sample ($n = 1,200$). Similarly, the study postulated that CAS-Arabic Edition is psychometrically appropriate for

this distinct cultural group of students by being culturally fair and less impacted by: (1) linguistic/translation issue; (2) economic/education disadvantaged positions; and (3) gender inequality.

Method

Participants

The CAS-Arabic Edition, consisting of 10 subtests, was administered to 49 Arab children (ages 5.0-7.11 years). The sample consisted of 21 males and 28 females, of whom 10 were 5.0 to 5.11 years old, 23 were 6.0 to 6.11 years old; and 16 were 7.0 to 7.11 years old. The mean age for females was 78.78; the mean age for males was 77.52. Research participants were recruited from the center of "MTLAM" in the city of Taibe. Currently the center serves 3,000 school age children and represents different social, religious, and economic strata. Subjects were provided with a parental consent form, an assent form, a letter of intent, and cover letter reviewing the importance and significance of the study to the Arabic-speaking community and to the development of such measures in Arabic, worldwide. All were chosen from the triangle area. Secondary data from the original standardization sample (were obtained from the original author of the CAS IQ test Dr. Naglieri) used to identify a matched sample of 49 American students on social demographic characteristics pertinent to the study; age, sex, and parental education attainment. Research participants ranged in age from 5.0 to 7.11 years, with a average age of approximately 78 months. The American sample was comprised of 60% Caucasians, 10% Hispanic and 30% African Americans, while all Arabic-speaking participants were of the same ethnicity—Palestinian. Twenty-nine percent of participants' parents had less than a high school education; 71% had more than four years of college.

Procedure

The data was collected by certified and trained teachers, all participants produced signed parental consent form, completed biographical data portion of the questionnaire. The biographical data collected was similar to that in the English form, all confidentiality and APA ethical rules when working with minors were followed. The administration of CAS-Arabic Edition-Basic Battery takes approximately 45 minutes. However, each participant was allotted a total of 60 minutes, allowing for time to discuss the forms and procedures with the parents, fill out the questionnaire, and establish rapport with the examiner.

Instrumentation

CAS is organized into three levels: full scale, PASS scale and, individual subtests.

The full scale is the overall measure of cognitive functioning and is composed of an equal weight composite of the four scales: planning, attention, simultaneous, and successive scales. The scale has a normative mean of 100 with standard deviation of 15. Each scale has a normative mean of 100 with standard deviation of 15. The makers of CAS allow two combinations of subtests to obtain the PASS scale and the full scale score: (1) the Basic Battery which includes eight subtests—two per each PASS scale (identified with an asterisk below); and (2) the Standard Battery which includes 12 subtests—three per each PASS scale. In the Standard Battery (12 subtests), each scale includes 3 subtests: CAS consists of four scales.

(1) Planning scale—Scale subtests include the subtests: (a) Matching numbers*; (b) Planned codes*; and (c) Planned connections;

(2) Attention scale—Subtests in this scale include (a) Expressive attention*; (b) Number detection*; and (c) Receptive attention;

(3) Simultaneous scale—This scale includes the following subtests: (a) Nonverbal matrices*; (b) Verbal-spatial relations* (verbally-loaded subtests); and (c) Figure memory—a non-verbal subtest;

(4) Successive scale—Three subtests comprise this scale: (a) a word series subtest (a subtest that is verbally loaded and examines memory of repeating series of words); (b) a sentence repetition subtest (contains little meaning and is structured in a way that aims to reduce the influence of simultaneous processing); and (c) a speech rate subtest.

CAS has very good psychometric properties; the internal reliability of the CAS standard scale ranges from a low of 0.95 to a high of 0.97. The scale reliability is more than sufficient (Bracken, 1987): Standard battery scales are 0.88 each for planning and attention, and 0.93 each for simultaneous and successive scales. The reliability coefficient for the basic battery is also as high; full scale ranges from 0.85 to 0.90 with average reliabilities for basic battery scales, 0.85. Subtest reliabilities range from 0.75 to 0.89 across subtests. The validity of CAS has been examined using content construct and criterion-related methods. CAS has also been validated through examination of its relationship with the Woodcock-Johnson Test of Achievement. The standard battery and the basic battery achieved high correlations of 0.73, and 0.74, respectively (Das-Naglieri, 1997). CAS has been shown to be effective for both diagnostic and remedial purposes, particularly with school-aged populations. The CAS-Arabic Edition test package included a recording form, response book, figure memory booklet, and a stimulus book. The recording form was used 100% of the time by the tester only. The response book included three subtests (matching numbers, planned codes, and number detections) to be completed by the examinee. The examinee was also required to complete "The figure memory" subtest. The stimulus book was controlled by the examiner and used to expose the test items to the examinee while the examiner records the examinee's responses. Scoring of performance on the CAS-Arabic Edition was done as in the original English version following the discontinuation rule was used were applies as well as time limit and time recording in seconds. More details on scoring are found in the Das-Naglieri CAS Administration Manual (1997).

Instrument Translation

The CAS was translated into Modern Standard Arabic—the native language for the population from which the research sample was selected. Guidelines recommended by Hambleton (1993) for cross cultural translation of assessment instruments were followed. Specifically the following steps were implemented: translation into target language, back translation to English and finally review by a professional who has not been involved in the first two tasks to compare the back-translated text to the original. Back-translation is the best known and most popular of the judgmental methods (Hambleton, 1993).

Structural and Administrative Modifications as Result of Pilot Test Findings

Pilot study was conducted to ensure smooth administration; results indicated that, the CAS-Arabic Edition should remain essentially true to the original English edition. Results suggested only minor changes that could be handled orally during testing. For example, the Cross sign in the "Verbal-Spatial" subtest was referred to as a "Plus Sign", and for the X, O in the planned codes subtest, we called X multiplication sign, and the "O" was called a "circle or an "O". As a result of the pilot study, it was decided to ask each research participant if he/she knew the signs, to identify their right hand and left hand, those subjects who failed were excused from the research sample. In addition, culturally biased symbols such as "Cross" used in the "Verbal-spatial subtest" were replaced by the "Plus" sign. Few of the pilot study participant's ages 5.0-5.11 failed to identify the picture

of a "Dinosaur", a decision was made to ask participants to identify and be informed of the name of the animal in the picture if they did not recognize the animal or did not name a different big animal, as did four children in the pilot sample.

The current study examined a number of variables that may influence the performance of Arabic children. Dependent variables in this study were CAS-Arabic Edition subtest scores, scaled scores, and the Full score. Independent variables included age, gender, parental educational attainment, and developmental and health status constituted the four independent variables of the study.

Data Collection

The planning scale (two subtests) and the attention scale (two subtests) were administered as recommended in the administration manual for the Basic Battery. The simultaneous scale (three subtests) and the successive scale (three subtests) were administered as recommended for use in the Standard Battery. Data were also collected from the administration of the two additional subtests that are not verbally loaded: figure memory (from the simultaneous scale) and speech rate (from the successive scale) in order to determine the impact of translation on performance. The full score of each examinee was derived from the basic battery test of eight subtests as recommended in the Basic Battery.

Data Analysis

In the current study, data was analyzed using SPSS 16.0. The study employed frequency distributions, means, and standard deviations to examine the quality and characteristics of the sample data. Subtest reliability coefficients were calculated using split-half methods for all subtests except Speech Rate. Cronbach's alpha was used to obtain reliability coefficients for all subscales within the CAS measure. For the Cronbach's alpha, because the two groups were matched, dependent t-test was used to determine differences in the means between the standardization sample (American) and the Arab sample. The mean, standard deviation, and standard errors of measurement, were used to determine the confidence intervals. T-test was used to determine if significant difference exists between matched samples. Independent t-test was used to compute if significant differences exist between the two independent groups (49 $vs.$ 1,200). Finally, a four-way ANOVA was computed to examine the interaction effect of the following variables: group (Arab $vs.$ American) parental education attainment, age, and gender.

Results

Internal Reliability

To measure internal consistency for the Basic Battery subtests, and the scales and full score of the CAS-Arabic Edition, the formula for the reliability of linear combinations (Nunnaly & Bernstein, 1994) was used. Reliability criterion is considered to be acceptable for tests of cognitive abilities at 0.80, subtests with reliabilities of 0.70-0.79 and are considered to be relatively reliable (Sattler & Hoge, 2004). Alpha at values of 0.90 or higher is considered to be extremely strong (Anastasi & Urbina, 1997). The reliability coefficient for the eight subtests that comprise the CAS Basic Battery-Arabic Edition ranged from 0.71 to 0.94. Average reliability coefficients were found in three subtests: verbal-spatial subtest (0.71), expressive attention (0.74), and matching number (0.79). Reliability coefficient for planning codes and expressive attention subtests were found to be at 0.84 and 0.83, respectively. The highest reliability coefficients were found in three subtests: number detection (0.92), word series (0.93), and sentence repetition (0.94).

The results of the reliability analysis for the PASS and full scales show that the four scales (i.e., planning, attention, simultaneous, and successive) exceeded the acceptable alpha level of 0.80. The reliability coefficients of the planning and simultaneous scales were at acceptable levels of 0.84 and 0.81, respectively, whereas the alpha coefficient for the attention scale and the successive scale were at the extremely strong levels of 0.91 and 0.94, respectively. Table 1 and Table 2 indicate that the full scale reliability coefficient was found to be at the strong level of 0.93.

Table 1

Internal Consistency Coefficients for the Performance

Subtests	Coefficient
Matching numbers	0.79
Planning codes	0.84
Expressive attention	0.74
Number detection	0.92
Speech rate	0.88
Non-verbal matrices	0.83
Verbal spatial	0.71
Figure memory	0.80
Word series	0.93
Sentence repetition	0.94

Note. Cronbach's alpha was used for the first five subtests and split-half method was used for the last five subtests. All estimates were based on $n = 49$.

Table 2

Internal Consistency Coefficients for the Performance on CAS-Arabic Edition by Scale

Scale	Name of subtests	Reliability coefficient
Planning scale	Matching numbers, planned codes	0.83
Attention	Expressive attention, number detection	0.91
Simultaneous	Nonverbal matrices, verbal-spatial relations	0.81
Successive	Word Series, sentence repetition	0.94
Full scale	All	0.93

Note. These reliability estimates were based on the method of linear combinations as implemented in the reliability procedure in SPSS. All estimates were based on $n = 49$.

Comparison between the reliability coefficients of the Arabic sample at the scale level with the American standardization sample shows close reliability coefficients in the planning scale (0.83 *vs.* 0.85), and higher reliability in the attention scale with the Arab sample (0.91) than with the American sample (0.84). The reliability coefficient of the simultaneous scale in the American sample (0.90) is higher than which was found in the Arab sample (0.81). The reliability of the successive scale for the Arab sample was found to be 0.94, while it was 0.90 for the American sample. The full scale reliability coefficient for the Arab sample was calculated at 0.93 *vs.* 0.87 with the American sample. Table 3 shows that the reliability coefficients based on the Arabic sample and the American sample are comparable with differences ranging from 0.02 to 0.09. Comparison of the reliability coefficients by scale for the American and the Arabic samples is presented in Table 3.

Further examination was carried out to examine whether translation had any potential impact on validity, in terms of sensitivity to linguistic or semantic adaptation from English version to Arabic version. The influence of language translation was evaluated with a four way ANOVA analysis of the three subtests that comprise the

simultaneous scale (i.e., nonverbal matrices, verbal-spatial, and figure memory) the three subtests of the Successive Scale (i.e., word series, sentence repetition, and speech rate). The results (Table 4) indicated no significant mean differences based on group, age, and gender. However, differences were found between the groups based on parental educational attainment level ($F(1, 590) = 17.44$, $p < 0.001$, MSE = 2994.70), suggesting that there is an overall effect of parental educational attainment on performance of the participants. There was also an interaction effect between the group and parental educational attainment ($F(1,590) = 3.95$, $p < 0.05$, MSE = 679.50), showing that the effect of parental education was not the same for both groups (American vs. Arab).

Table 3

Internal Reliability Coefficients for the CAS Scales: Arab Sample vs. Standardization Sample

Scale	Name of subtests	CAS-Arabic Edition	CAS-English Edition
Planning scale	Matching numbers, planned codes	0.83	0.85
Attention scale	Expressive attention, number detection	0.91	0.84
Simultaneous scale	Nonverbal matrices, verbal spatial relations	0.81	0.90
Successive scale	Word series, sentence repetition	0.94	0.90
Full scale-Basic Battery	All eight subtests	0.93	0.87

Notes. These reliability estimates were based on the method of linear combinations as implemented in the reliability procedure in SPSS. All estimates were based on Arab sample $n = 49$, American sample for children < 8 years old $n = 1,200$.

Table 4

Summary of Means of the Two Groups and Parental Education Attainment level

Group/Parental education attainment	American	Arabs	Total
Less than high school	101.53 (13.73) $n = 271$	90.71 (12.58) $n = 14$	101.00 (13.86) $n = 285$
College and over	106.45 (13.7) $n = 295$	106.35 (16.88) $n = 34$	106.44 (12.81) $n = 329$
Total	104.09 (13.22) $n = 566$	101.79 (17.19) $n = 48$	

Note. Means (standard deviation) and number of subjects.

Two-way ANOVA analysis show significant difference between the three subtests in the successive scale ($F(2,2396) = 3.68$, MSE = 5.15, $p < 0.05$). In order to identify the source of the difference, an additional analysis was conducted which demonstrated higher mean scores in the word series subtest (11.3) as compared to the other two subtests (10.2 and 10.3) for sentence repetition and speech rate, respectively ($F(1,1198) = 3.457$, MSE = 5.91, $p < 0.0001$). In addition, the interaction between subscale and group was significant ($F(2,2396) = 4.27$, MSE = 5.15, $p < 0.05$). Additional analysis showed that the source of the interaction was the difference between subtests within the Arab group ($F(2,96) = 3.70$, MSE = 5.69, $p < 0.05$); the mean differences between the subtests within the American group were not significant. Comparison of the means per subtest between the two groups is presented in Table 5.

To determine if CAS-Arabic Edition is psychometrically appropriate for this distinct cultural group, the study hypothesized that the performance of the two groups (Arab and American) are comparable. A matching design was used, and the Arab participants were matched with participants from the original CAS standardization sample on the variables of age, gender, and parental education. For the age variable, three age groups were created (5.0-5.11, 6.0-6.11, and 7.0-7.11). The tables below present the means (Table 6, Table 7 Table 8), standard deviations, and the results of the *t*-tests for each subtest, PASS scale, and for the full scale.

No significant differences between the groups were detected among any of the subtests, the PASS scales, or the full scale. Means and standard deviations for the PASS scales, as well as the full scale are presented in Table 12. Table 13 presents the paired differences between the groups per subtest. Table 6 presents the paired differences between the two groups for the PASS scales and the full scale.

Table 5

Subtests Means and Standard Deviation for the Matched Groups

Subtests	Arab		American	
	Mean	Std	Mean	Std
Matching number	10.61	2.55	9.87	3.04
Planned codes	10.75	6.25	10.79	3.15
Non-verbal matrices	10.08	2.69	10.26	3.11
Verbal-spatial	9.48	3.06	10.10	2.91
Figure memory	10.46	3.27	10.00	3.05
Expressive attention	10.97	2.39	10.26	3.54
Number detection	10.04	2.18	9.93	2.54
Word series	11.34	3.39	10.34	3.37
Sentence repetition	10.16	2.04	11.02	2.78
Speech rate	10.18	2.29	10.68	2.29

Note. All results were based on $n = 49$ for each group except: planned codes $n = 48$, verbal spatial $n = 46$.

Table 6

Means and Standard Deviation by Scale and Full Scale

Scale and full scale	Arab		American	
	Mean	Std	Mean	Std
Planning	100.41	14.27	101.85	15.69
Simultaneous	98.79	13.45	100.95	13.45
Successive	100.85	13.33	103.85	15.40
Attention	102.86	11.81	101.15	14.05
Full scale	101.02	11.47	102.71	13.87

Note. All results were based on planning $n = 48$, simultaneous $n = 49$, successive $n = 49$, attention $n = 46$, full scale $n = 45$.

Table 7

Paired Differences per Subtest Between Arab and American Groups

Subtests Arab-Americans	Means	Std. Deviation	Std. Error Mean	95% Confidence interval of the difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Matching numbers	0.73	3.63	0.52	-0.30	1.77	1.41	48	0.16
Planned codes	-0.04	7.07	1.02	-2.09	2.01	-0.04	47	0.97
Nonverbal matrices	-0.18	3.30	0.47	-1.13	0.76	-0.39	48	0.70
Verbal-spatial relations	-0.61	3.98	0.57	-1.75	0.53	-1.07	48	0.29
Figure memory	0.47	4.03	0.57	-0.68	1.62	0.81	48	0.42
Expressive attention	0.71	3.78	0.54	-0.37	1.80	1.32	48	0.19
Number detection	0.11	3.36	0.49	-0.88	1.10	0.22	45	0.83
Word series	1.0	4.75	0.68	-0.36	2.36	1.47	48	0.15
Sentence repetition	-0.86	3.37	0.48	-1.82	0.11	-1.77	48	0.08
Speech rate	-0.50	3.71	0.53	-1.57	0.57	-0.93	47	0.36

Note. None of the tests were significant; consequently effect sizes were not reported.

Table 8

Paired Differences for PASS Scale and Full Scale Between Arab and American Groups

PASS scales Arab-Americans	Mean	Std. Deviation	Std. Error Mean	95% Confidence interval of the difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Planning scale	-1.43	18.13	2.61	-6.70	3.82	-0.55	47	0.58
Simultaneous	-2.16	16.36	2.33	-6.86	2.53	-0.92	48	0.36
Attention	1.71	17.85	2.63	-3.58	7.01	0.65	45	0.52
Successive	-3.00	20.40	2.91	-8.86	2.86	-1.02	48	0.31
Full basic scale	-1.69	16.79	2.50	-6.73	3.35	-0.67	44	0.50

Note. All estimates are based in n = 45-49 depending on the scale.

Since comparison between the scores of the Arabic sample (n = 49) and the total original CAS standardization sample (n = 1,200) would provide more reliable conclusions about the validity of CAS-Arabic Edition, more analysis was conducted with respect to the following: (1) individual subtests; and (2) the four scales and the full IQ score also support a high degree of validity. Our fourth hypothesis, thus, was partially supported, independent t-test analysis (see Table 5) yield significant differences between the groups in two subtests: (1) the expressive attention subtest (t(1226) = 1.99, p < 0.05), in which the effect size was calculated as 0.28 using the differences between the two means of the two groups divided by the standard deviation of the American sample; and (2) the word series subtest (t(1247) = 2.8, p < 0.005) in which the effect size was calculated as 0.40 using the same formula.

Table 9

Subtest Scores Arab Sample vs. American Sample

	Arab			American		
	n	Mean	Std Deviation	n	Mean	Std Deviation
Matching number	49	10.61	2.55	1,190	9.97	3.02
Planned codes	49	10.75	6.19	1,157	10.08	3.00
Non-verbal matrices	49	10.08	2.69	1,199	10.04	2.95
Verbal spatial	49	9.48	3.06	1,200	10.04	2.85
Figure memory	49	10.46	3.27	1,196	10.17	2.94
Expressive attention	49	10.97	2.39	1,179	10.09	3.07
Number detection	49	9.89	2.21	1,169	10.03	2.91
Word series	49	11.34	3.39	1,200	10.05	3.17
Sentence repetition	49	10.16	2.04	1,200	10.10	2.98
Speech rate	49	10.26	2.33	1,151	10.11	3.04

Analysis indicated no significant difference between the performance of the groups on the PASS scales and the full scale. Table 11 presents the results of the analysis for the Arab sample (n = 49) and the American sample (n = 1,127) for subtests. Table 12 presents results for the PASS scale and the Full Scale levels and Table 18 provides a comparison of the means and standard deviations for the groups by subtest.

Finally, the study postulated that the validity of CAS-Arabic Edition will not be effected by gender inequality, ethnicity, or educational disadvantages. A four-way ANOVA (analysis of variance) was conducted using the following independent factors: group (Arab *vs.* American), age, gender, and parental education attainment; and the following dependent variables: subtests, scales, and full scale score. Results showed that significance differences exist for parental educational attainment and education by group interaction. In other

words, parental education had an overall or main effect on the performance of both groups; however, the effect of parental educational attainment was not the same for the two groups. A clear advantage was found for the Arab group when comparing the mean performance of high school educated parents with that of college educated parents. Figure 1 presents the effect of parental education level among Arab participants on performance by subtest.

Table 10

Subtest scores Arab Sample vs. American Sample

	Arab			American		
	n	Mean	Std Deviation	*n*	Mean	Std Deviation
Matching number	49	10.61	2.55	1,190	9.97	3.02
Planned codes	49	10.75	6.19	1,157	10.08	3.00
Non-verbal matrices	49	10.08	2.69	1,199	10.04	2.95
Verbal spatial	49	9.48	3.06	1,200	10.04	2.85
Figure memory	49	10.46	3.27	1,196	10.17	2.94
Expressive attention	49	10.97	2.39	1,179	10.09	3.07
Number detection	49	9.89	2.21	1,169	10.03	2.91
Word series	49	11.34	3.39	1,200	10.05	3.17
Sentence repetition	49	10.16	2.04	1,200	10.10	2.98
Speech rate	49	10.26	2.33	1,151	10.11	3.04

Table 11

PASS Scale and Full Scale scores Arab Sample vs. American Sample

	Arab			American		
	n	Mean	Std Deviation	*n*	Mean	Std Deviation
Planning scale	49	100.71	14.27	1,156	100.18	15.60
Simultaneous scale	49	98.79	13.45	1,199	100.12	13.99
Attention scale	49	102.14	12.00	1,160	100.50	14.37
Successive scale	49	100.85	13.33	1,199	100.42	15.16
Full scale score	49	100.91	11.31	1,127	100.56	14.70

Note. $^{*}p < 0.05$; $^{**}p < 0.01$.

Table 12

Comparison of the Means and Standard Deviation of the Two Groups by Subtest

	CAS Arabic Edition *n* = 49				CAS English Edition *n* = 1,200				Means Differences
	n	Mean	Standard Deviation	Std Error Mean	*n*	Mean	Standard Deviation	Std Error Mean	
Matching numbers	49	10.61	2.55	0.36	1,190	9.97	3.02	0.087	0.63
Planned codes	49	10.75	6.19	0.88	1,157	10.08	3.00	0.088	0.67
Nonverbal matrices	49	10.08	2.69	0.38	1,199	10.04	2.95	0.852	0.88
Verbal-spatial	49	9.48	3.06	0.43	1,201	10.04	2.85	0.082	-0.13
Figure memory	49	10.46	3.27	0.47	1,196	10.17	2.94	0.085	0.29
Expressive attention	49	10.97	2.39	0.34	1,179	10.09	3.07	0.089	0.18
Number detection	49	9.89	2.21	0.31	1,169	10.03	2.91	0.085	-0.55
Word series	49	11.34	3.39	0.48	1,200	10.05	3.17	0.091	0.29
Sentence repetition	49	10.16	2.04	0.29	1,200	10.10	2.98	0.086	0.06
Speech rate	49	10.26	2.33	0.33	1,151	10.11	3.04	0.089	1.29

Table 12 to be continued

	Levene's test for equality of variances		t-test for equality of means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Matching numbers	3.67	0.06	1.45	1,237	0.14	0.63	0.43	-0.22	1.49
Planned codes	2.37	0.12	1.43	1,204	0.15	0.66	0.46	-0.24	1.58
Non-verbal matrices	0.82	0.36	0.09	1,246	0.92	0.04	0.42	-0.79	0.88
Visual-spatial relations	1.11	0.29	-1.33	1,248	0.18	-0.55	0.41	-1.37	0.26
Figure memory	1.05	0.30	0.68	1,243	0.49	0.29	0.43	-0.54	1.14
Expressive attention	1.77	0.18	1.98*	1,226	0.047	0.88	0.44	0.011	1.75
Number detection	2.92	0.09	-0.31	1,216	0.75	-0.13	0.42	-0.95	0.69
Word series	0.08	0.78	2.79**	1,247	0.005	1.29	0.46	0.38	2.20
Sentence repetition	7.16	0.008***	0.20	1,247	0.83	0.06	0.30	-0.55	0.67
Speech rate	4.92	0.03***	0.43	1,198	0.66	0.15	0.34	-0.54	0.84
Planning scale	0.88	0.35	0.23	1,203	0.81	0.53	2.26	-3.91	4.98
Simultaneous scale	0.38	0.54	-0.65	1,246	0.51	-1.33	2.03	-5.32	2.66
Attention scale	1.68	0.19	0.78	1,207	0.43	1.64	2.08	-2.44	5.72
Successive scale	1.69	0.19	0.19	1,246	0.84	0.43	2.20	-3.88	4.74
Full scale	3.76	0.05	0.16	1,174	0.86	0.35	2.12	-3.82	4.53

Notes. None of the Levine tests conducted were significant so the equality of variance should be assumed; * $p < 0.05$, ** $p < 0.01$; Means comparisons show no significance; Consequently no effect sizes were reported; ***Equality of variances is not assumed for sentence repetition and speech rate.
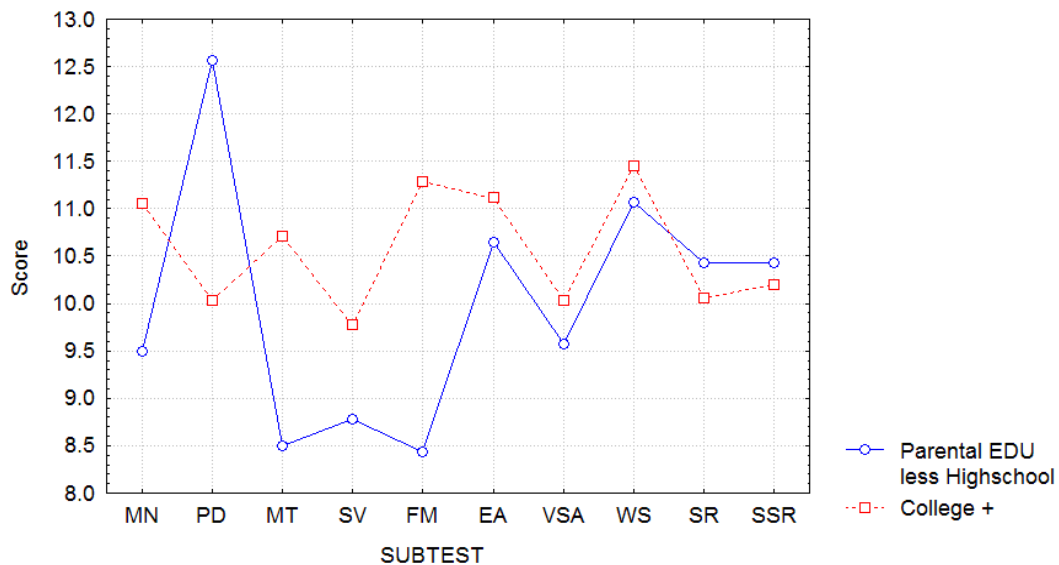

Figure11. Arab sample performance when controlling for parental education level (*Notes.* MN = Matching Numbers; PD = Planned Codes; MT = Nonverbal Matrices; SV = Verbal-Spatial Relations; FM = Figure Memory; EA = Expressive Memory; VSA = Number Detection; WS = Word Series; SR = Sentence Repetition; SSR = Speech Rate).

The performance on the full scale indicated that Arab students with college-educated parents scored sixteen points more than those whose parents had less than a high school education. However, American students with college educated parents had only a five point advantage over those whose parents had less than high school (see Table 13).

Table 13

Education Means and Standard Deviation by Group

| | Arab | | American | |
	Means	Std	Means	Std
Less than high school	90.70	12.58	101.53	13.73
College	106.35	16.88	106.45	12.30

Notes. Based on: Arab sample: Less than high school $n = 14$; College $n = 34$; American sample: Less than high school $n = 271$, College $n = 295$.

Discussion

Despite their clinical and educational utility, cognitive assessment tools remain under-developed, under-researched, and under-utilized in non-Western societies (Ahmed & Geilen, 1998). An examination of extant literature in the Arab world, outside of Israel, reveals that research on standardized assessment of cognitive abilities is virtually nonexistent in Middle-Eastern, Arabic-speaking societies. Furthermore, most work on the development of standardized psychological tests in Arab societies is outdated and has heavily relied on translated tests from Western societies, most notably the United States (Ahmed & Geilen, 1998). The CAS-Arabic Edition was developed from the Das-Naglieri CAS. The fact that language translation did not seem to impact performance on the Arabic Edition of the CAS contributes to the construct validity of CAS. Criterion validity for the CAS has also been established through numerous studies and correlations with results from other intelligence tests by the test developer and others. It is not yet possible to validate the findings of CAS-Arabic Edition in this way with other intelligence tests in Arabic because of the absence and lack of up-to-date valid measures in that language. However, it is worth noting that an intelligence test result is only one component of the diagnostic process. As such, it must not be the sole tool for diagnosis, but should be used as a screening tool. Reliability coefficients findings on the subtest level for the CAS-Arabic Edition were found to be acceptable, ranging from 0.71 to 0.94 compared to the CAS-English Edition subtest reliability coefficients ranging from 0.75-0.89. Similarly, high positive reliability coefficients were found at the PASS scale level, ranging from 0.81 to as high as 0.94. The total full Basic Battery scale reliability coefficient was 0.93. These results are consistent with the results obtained by the Naglieri and Das standardization sample on all three levels: subtest, scale, and full scale score (Naglieri & Das, 1997).

Translation may have had some impact on Arab participants' performance on the successive processing scale. A significant difference between the three subtests of the successive scale was found. The successive scale is a highly verbal scale; it includes word series, sentence repetition, and speech rate. The significant differences that exist between the subtests are due to the differences in the means of the three subtests. While the American group showed homogeneity in mean scores across the three subtests: word series (10.34), sentence repetition (11.02), and speech rate (10.68), the Arab sample obtained a higher mean in the word series subtest (11.34) compared to the other two subtests: sentence repetition (10.16) and speech rate (10.18). One plausible explanation is that the Arabic language is poetic in nature; this could explain the high scores in the Word Series Subtest. It also may explain lower scores in the other two subtests, particularly the sentence repetition subtest. The sentence repetition subtest consists of 20 sentences that are read to the child. Each sentence is composed of color words (for example, "The blue is yellowing"), and the child is required to repeat the sentence as it was presented. "These color words are used so that the sentence contains little meaning and help reduces the influence of simultaneous processing. Successful completion of the task demands the

appreciation of the sentence syntax" (Naglieri, 1999). Further analysis and research is needed in this field by psycholinguists.

The two-tailed *t*-test analysis demonstrated no significant differences between the Arab sample and the standardization sample on any of the subtests, nor on the PASS scales or the full scale. This suggests that an intelligence measure driven by psychological processes is more likely to produce similar results when samples are matched on core variables such as age, gender, and educational attainment of the parents.

Comparison of the performance of the Arab sample ($n = 49$) with the entire standardization sample ($n = 1,200$) with respect to individual subtests, PASS scales, and full scale scores, results indicated no significant differences were found between the groups for most subtests, nor for the PASS scales or the full scale. However, significant difference was found between the two groups on two subtests: expressive attention and word series. Expressive attention is one of two attention scale subtests designed to measure selectivity and ability to shift attention. The expressive attention subtest establishes an interference condition after administration of items without interferences. The test requires the child to avoid responding to the visual feature (size of animal in a picture) and to respond to another feature (i.e., actual size of the animal in reality). The expressive attention subtest examines selective attention, the ability to ignore the physical attention of the visual stimuli, to mentally compare the meaning of another dimension of the stimuli, and finally to give one of two competing possible answers (In English, the required response is "small" or "big"; in Arabic it is "Sgeer"—small, or "Kbeer"—big). Competition between very similar responses makes the task more difficult (Kornblum, 1973). Data analysis confirmed this expectation and indicated no significant age-related differences on subtests, PASS scales or on the Full Scores. The analysis also show no differences exist between the genders ($f = 28$, $m = 21$) on the performance of the Arabic-speaking participants on any of the subtests. However, males performed better on the PASS scale level (mean = 103.0) than did females (mean = 97.0). Analysis shows no interaction was found between gender and performance on the full scale level. One possible explanation for the difference between male-female performance on the PASS scale level could be that in traditional societies, boys are more exposed to outside stimuli than girls are, due to gender roles in such societies. Girls are generally are more restricted to the home than boys. In addition, there was a significant effect of parental educational attainment on subtest performance. The performance of children with parents of college educational level was significantly higher than those with parental education at less than high school level. One explanation for this finding may be the influence of the immediate environment on performance. The analysis also shows a significant effect of parental educational attainment on children's performance on the scale level. Arab children of parents with college education performed at mean = 106.35 (std 16.88), American at mean =106.45 (std 12.30). Arab children with parent educational attainment of less than high school achieved a mean score of 90.70 (std = 12.58) compared to American children of parents with less than high school education, who achieved a mean of 101.53 (std = 13.73). No effect of parental educational attainment was found on the full Scale level for Arab children. These findings confirm the results of studies on the impact of parental education on performance of children (Benjamin, 1993; Desforges & Abouchaar, 2003; Stenberg, Lamborn, Dornbusch, & Darling, 1992). However, the findings are limited in their generalizability and may be attributed to the small number of participants whose parents had less than high school education ($n = 14$). A floor effect may have been generated for this group, while a ceiling effect operated for the group with college educated parents.

Additional research regarding the CAS-Arabic Edition is warranted, a bigger representative sample is required and administration of test to older population (8.0-17.11) is warranted. Yet, the findings of this study

confirms that psychological process-driven measures are as valid and as reliable as traditional measures, yet may be more culturally fair and thus more adaptable across cultures and languages. Results of the study support this premise and indicate that this measure is reliable and valid measures that can be adopted to meet the needs of Arabic speaking students. Results indicate that the reliability of the CAS-Arabic Edition is comparable to that of the CAS used with the United States sample. It is also has considerable support for its validity with Arabic-speaking children in Israel ages of 5.0-7.11.

References

Ahmed, R. A., & Gielen, U. P. (1998). *Psychology in the Arab countries*. Menoufia, Egypt: Menoufia University Press.

Ahmed, R. A. (1998). Bibliography of psychological studies in the Arab world (unpublished, available from the author).

Anastasi, A. (1988). *Psychological testing*. New York: Macmillan Publishing Company.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice-Hall International.

APA. (1994). Guidelines for child custody evaluations in divorce proceedings. *American Psychologist, 49*, 677-680.

Benjamin, A. L. (1993). *Parents' literacy and their children's success in school: Recent research, promising practices and research implications*. Education Research Report.

Beller, M. (1992). *Admissions procedures in Israeli universities: Psychometricand social considerations*. Research Report. Jerusalem: National Institute for Testing and Evaluation.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*, 313-326.

Bracken, B. A., & McCallum, R. S. (1999). International testing matters: The Universal Nonverbal Intelligence Test (UNIT). *The International Test Commission Newsletter, 9*(1),7-12.

Cahan, S., & Noyman, A. (2001). The K-ABC mental processing scale: A valid measure of "pure" intelligence? *Educational and Psychological Measurement, 61*, 827-840.

Ceci, S., & Williams, W. (2000). *The nature-nurture debate*. Blackwell Publishing.

Coursen-Neff, Z. (2005). Discrimination against Palestinian Arab children in the Israeli educational system. *International Law and Politics, 36*, 749-816.

Das, J. P., & Naglieri, J. A. (1997). *Naglieri-Das cognitive assessment system*. Chicago: Riverside.

Desforges, C., & Abouchaar, A. (2003). *The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review*. London: Department for Education and Skills.

Feuerstein, R., Randy, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment*. Baltimore: University Park Press.

Hambleton, R. K. (1993). Translating achievements tests for use in cross-national studies. *European Journal of Psychological Assessment, 9*, 54-65.

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQL instruments: The universalist approach. *Qual Life Res 1998, 7*, 323-335.

Hunt, D., & Randhawa, B. S. (1983). Cognitive processes and achievement. *Alberta Journal of Educational Research, 29*, 206-215.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12*(1), 123-129.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York: Russell Age Foundation.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12*(1), 123-129.

Hunt, E., Frost, N., & Lunneborg, C. (1995). Individual difference in cognition: New approach to intelligence. In G. H. Bower (Ed.), *The psychology of learning and motivation*.

Khaleefa, O., & Ashria, I. (1995). Intelligence testing in an Afro-Arab Islamic culture: The Northern Sudan. *Journal of Islamic Studies, 6*, 222-233.

Lieblich, A. (1975). The predictive validity of the WPPSI with Israeli children. *Educational and Psychological Measurement, 35*, 473-475.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system: Interpretive handbook*. Chicago: Riverside Publishing Company.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.

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

Sattler, J., & Hoge, R. D. (2004). *Assessment of children: Behavioral, social and clinical foundations* (5th ed.). Sattler Publishing.

Scott, G. (1950). Measuring sudanese intelligence. *The British Journal of Educational Psychology, 20*, 43-54.

Stenberg, L., Lamborn, S., Dornbusch, S. M., & Darling, N. (1992). Impact of parenting practices on adolescent achievement: Authoritative parenting, school involvement, and encouragement to succeed. *Child Development, 63*(5), 1266-1281.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1996) *Successful intelligence*. New York: Simon and Schuster.

Sternberg, R. J. (1999). The theory of successful intelligence. *Rev. Gen. Psychol., 3*, 292-316.

Sternberg, R. J. (Ed.). (2004). *International handbook of intelligence*. New York:Cambridge University Press.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49*, 1134-1139.

Woodcock-Johnson. (1989). *Woodcock-Johnson tests of cognitive ability-revised*. Itasca, I.L.: Riverside Publishing.

Zeidner, M. (2004). In Sternberg, R. J. (Ed.). *The Israeli intelligence experience. International handbook of intelligence*. New York: Cambridge University Press.