Behavioral Psychology / Psicología Conductual, Vol. 24, Nº 3, 2016, pp. 531-554

CROSS-CULTURAL ADAPTATION OF EXECUTIVE FUNCTION TESTS FOR ASSESSMENTS OF TRAUMATIC BRAIN INJURY PATIENTS IN SOUTHEAST IRAN

Heshmatollah Ghawami[1], Mahvash Raghibi[1], Bahman Kord Tamini[1], Behrooz Dolatshahi[2], and Vafa Rahimi-Movaghar[3]
[1]University of Sistan and Baluchestan; [2]University of Social Welfare and Rehabilitation Sciences; [3]Tehran University of Medical Sciences (Iran)

Abstract

The Delis-Kaplan Executive Function System (D-KEFS) is a "greatest hits" collection of commonly used tests of executive functioning. We mainly aimed to develop a Persian version of D-KEFS for brain-damaged patients in Zahedan, Iran, and to provide preliminary validation evidence, applying a normal sample with demographic traits of the patients. In the judgmental phase, we conducted several local field studies and non-standard pilot administrations, and accordingly we made various cross-cultural adaptations. For the statistical phase, the provided materials along with four tests from the Behavioral Assessment of the Dysexecutive Syndrome (BADS) were administered to 75 healthy individuals. Within 12 to 30 days, 24 subjects were retested. Relatively high reliabilities were obtained for most of the D-KEFS measures. In validity analysis, strong correlations were found among the majority of the scores within the tests; correlations between various D-KEFS tests were in the range of weak to moderate; and significant correlations were found between the majority of D-KEFS executive scores and BADS scores. In conclusion, the adapted tests show acceptable psychometric properties in assessing the complex, multidimensional construct of executive functioning.

KEY WORDS: *test adaptation, validation, D-KEFS, executive functions, traumatic brain injury*.

Resumen

El "Sistema Delis-Kaplan de la función ejecutiva" (*Delis-Kaplan Executive Function System;* D-KEFS) es una colección de "grandes éxitos" de tests utilizados para la evaluación del funcionamiento ejecutivo. Nuestro objetivo principal fue desarrollar una versión persa del D-KEFS para pacientes con daño cerebral en

This research was supported in part by Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences and the University of Sistan and Baluchestan. The authors would like to thank Prof. Vicki Anderson for her advice in selecting the executive tests, and also to Dr. Mahdi Mohammadi, Dr. Ismaiel Noormohammadi, Dr. Yahya Kaikhaie, and Ms. Somayeh Okati, for their valuable helps. We are grateful to Dr. Mehrnaz Narouei-Nejad and Dr. Mohammad-Reza Rohani for their financial assistance.

Correspondence: Vafa Rahimi-Movaghar, Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences, Sina Hospital, Hassan-Abad Sq., Imam Khomeini St., Tehran, 11365-3876 (Iran). E-mail: v_rahimi@sina.tums.ac.ir

Zahedan, Irán, y proporcionar pruebas preliminares de su validación al aplicarlo a una muestra de pacientes locales. En la fase de evaluación, llevamos a cabo varios estudios de campo y aplicaciones piloto no estándar y, en consonancia, realizamos varias adaptaciones transculturales. En la fase estadística, dichos materiales junto con cuatro tests de la "Evaluación conductual del síndrome disejecutivo" *(Behavioral Assessment of the Dysexecutive Syndrome;* BADS) se administraron a 75 individuos sanos. Entre 12 y 30 días después, se volvió a evaluar (retest) a 24 sujetos. Se obtuvieron fiabilidades relativamente elevadas para la mayoría de las medidas del D-KEFS. En los análisis de validación, se encontraron elevadas correlaciones entre la mayoría de las puntuaciones intratests. Las correlaciones entre varios de los tests del D-KEFS se hallaban en el rango de débiles a moderadas y se encontraron correlaciones significativas entre la mayoría de las puntuaciones ejecutivas del D-KEFS y de la BADS. En conclusión, los tests adaptados muestran propiedades psicométricas aceptables para evaluar el complejo constructo multidimensional del funcionamiento ejecutivo.
PALABRAS CLAVE: *adaptación de test, validación, D-KEFS, funciones ejecutivas, lesiones cerebrales traumáticas.*

Introduction

Given the existence of thousands of different cultures around the world and the acknowledged, substantial impact of culture on the brain organization of cognition, during recent decades, the traditional westernized trend in neuropsychological assessment, in which the content and standardization of cognitive tests have been focused mostly on the western, white, English-speaking, middle-class, and educated cultural populations, has encountered serious criticisms (for more discussion, see Ardila, 1995, 2007; Nell, 2000; Van de Vijver & Leung, 1997). Ardila (1995) suggested some key directions to cross-cultural research, including the standardization of current neuropsychological tests in different cultural contexts and the development of new neuropsychological instruments, appropriate for different cultural groups; "neuropsychological tests must be adapted (i.e., redeveloped; not just translated) when applied to a cultural group different from the group in which they were originally developed" (p. 148).

Iran is among the five countries with the world's highest rates of road traffic crashes (RTCs) that, in addition to the deaths, cause 685,611 non-fatal injuries annually (World Health Organization, 2009, 2013). In 2005, over one million Iranians were injured (non-fatally) due to RTCs, in which the rate of the injuries peaked in the age group of 15-24 years (Bhalla, Sharaz, Abraham, Bartels, & Yeh, 2011). Most of the injured patients sustain traumatic brain injuries (TBIs) and ensuing cognitive complications, especially, chronic executive deficits resulting in long-term everyday dysfunctions. Despite the urgent need for valid neuropsychological tools in the intervention and rehabilitation programs of the patients, there are very few tests appropriately adapted or developed for assessments of Iranian clinical and non-clinical populations. This paper reports our efforts to adapt and preliminarily validate one of the widely used, western executive batteries for a group of TBI patients in the southeast of Iran.

Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001a) is a relatively comprehensive, US-standardized set of tests applied mainly to measure cognitive (*cold*) executive functions of the brain in both verbal and nonverbal modalities. The battery is a "greatest hits" collection of the commonly used executive tests (Lezak, Howieson, Bigler, & Tranel, 2012), composed of nine stand-alone tests co-normed on a stratified U.S. sample of 1,750 individuals aged 8 to 89 years (Delis et al., 2001c). The tests are Trail Making Test (TMT), Verbal Fluency Test (VFT), Design Fluency Test (DFT), Color-Word Interference Test (CWIT), Sorting Test (ST), Twenty Questions Test (TQT), Word Context Test (WCT), Tower Test (TT), and Proverb Test (PT).

In addition to the provision of a wide variety of multi-level normative scores, the D-KEFS has some other new features, making it seemingly more desirable for the assessments of TBI patients, especially for mild brain damages: (a) It applies a cognitive-process approach (Kaplan, 1988), parceling the contributions of more fundamental cognitive skills out from the multiple higher-level cognitive functions, thereby allowing to better identify the neurocognitive mechanisms underlying deficient performance; (b) several switching conditions, capture stimuli, procedures for increasing the processing demands, and higher ceiling and lower floors were added in order to enhance the sensitivity; and (c) the tests are in a game-like format without providing right/wrong feedbacks and are hence less frustrating for brain damaged patients who struggle more on the tests.

Delis et al. (2001c) pointed out that the validity of most of the tests (modified tests) was previously proven in the scientific literature during the past 50 years or more. In the manual, in addition to the preliminary evidence of convergent, discriminant, and clinical validities, Delis and colleagues also reported intratest and interest correlations between various measures of D-KEFS. Since the publication, D-KEFS package has generally been reviewed positively (e.g., Homack, Lee & Riccio, 2005; Shunk, Davis, & Dean, 2006; Swanson, 2005), although some reviewers have been critical, particularly, of the presented evidence of reliability and validity in the manual (e.g., Baron, 2004; Schmidt, 2003; Strauss, Sherman, & Spreen, 2006). Crawford, Sutherland, and Garthwaite (2008) estimated reliability coefficients for the contrast measures reported in the D-KEFS manual (Delis et al., 2001c). They obtained low reliabilities for the majority of the contrast measures, concluding that these measures are not appropriate for neuropsychological decision making. Nowadays, among the psychometric strengths of D-KEFS is a considerable body of literature establishing several dimensions of validity for the tests: convergent validity (e.g., Floyd et al., 2006); divergent/discriminant validity (e.g., Davis, Pierson, & Finch, 2011); criterion-ecological validity (e.g., Altemeier, Jones, Abbott, & Berninger, 2006; Jefferson, Paul, Ozonoff, & Cohen, 2006; Mitchell & Miller, 2008); criterion-retrospective/postdictive validity (e.g., Hancock, Tapscott, & Hoaken, 2010); also, the tests have measured executive dysfunctions in a wide range of clinical populations including patients with frontal lobe lesions (e.g., Lovstad et al., 2012; Yochim, Baldo, Kane, & Delis, 2009; for further clinical validity reviews, see also Delis, Kramer, Kaplan and Holdnack, 2004; Fine & Delis, 2011; Jones-Chesters, 2008).

The current study aimed to develop, for the first time, an appropriate Persian version of the D-KEFS tests for assessments of brain-damaged patients in Zahedan, Iran and to provide preliminary validation evidence, applying a normal sample of Baluch bilingual youth with demographic traits of the patients. The utility of the tests with TBI patients is reported in another study by Ghawami, Sadeghi, Raghibi, and Rahimi-Movaghar (2016).

Method

Phase 1 - A priori procedures: cross-cultural adaptation and content validation

Target population. Our target population was a normal group of Iranian bilingual Baluch men with demographic traits of brain-damaged patients admitted to Khatam-ol-anbiah Hospital, the main regional trauma center, in Zahedan, the capital of Sistan and Baluchestan province, in the southeast of Iran. According to the TBI registry data during 2011-2013 as well as the previous reports (e.g., Ansari-Moghaddam et al., 2012; Chardoli & Rahimi-Movaghar, 2006), the supermajority of the patients are young Baluch bilingual men, aged 15 to 40 years (peaked around 25), with lower-level education (0 to 12 years). The province is a poor socioeconomic region, where most youth of the target population drop out of school and engage in illicit cross-border trade activities (goods, drugs, fuel, etc.). Baluch (or Baloch) people are an ethnic group of tribes inhabiting mainly neighboring areas of Iran, Pakistan and Afghanistan; the population is mostly religious Sunni Muslims, bearing religion as the main part of their culture. Iranian Baluchs learn and use only Persian, as the formal language of Iran at school and work; even their religious leaders deliver the sermons in Persian. They use Baluchi language only in oral form at home; our participants had no knowledge of its written form. Therefore, given the written format of parts of the D-KEFS tests, we used Persian as the language of administration in this study.

Test adaptation procedure. In transferring cognitive tests to the non-western cultural and linguistic contexts, the extent and type of the changes depends on several factors such as the nature of tests and the cultural/linguistic distance between the source and target populations. For example, for "cross-culturally identical domains" (Poortinga, 1989), sometimes only the test instructions need to be translated whereas, for high cultural loading tests with idiomatic expressions (e.g., the proverbs), most of the items need to be changed. Thus, our adaptation of the D-KEFS tests on the adoption-assembly continuum (Malda et al., 2008) varied for different tests: collecting new items for more culture-bound tests (e.g., for the WCT and PT), changing only some parts/conditions of the test (e.g., for the TMT), or translating only the test instructions (e.g., for the TT). In addition to the initial aggregate pilot administration, an iterative combination of local field studies for selecting new items, followed by non-standard pilot administrations (Van de Vijver & Hambleton, 1996; Van de Vijver & Leung, 1997; Van de Vijver & Tanzer, 2004) were conducted for various D-KEFS tests. Depending on the requirements of different tests, various theory-, language-, culture-, and familiarity/recognizability-driven adaptations (Malda et al., 2008) were applied. Adopting a judgmental

committee approach (Van de Vijver & Leung, 1997; Van de Vijver & Tanzer, 2004), all steps of the adaptation process were under advisement, supervision, and content judgment of a mixed committee with different areas of expertise. The committee members, who were all PhD holders and had related research experiences on the Baluchi population comprised of two native trilingual experts (proficient in Baluchi, Persian, and English; one with expertise in English literature and psycholinguistics, and the other one in psychology), and three bilingual experts (proficient in both Persian and English; one with expertise in psychometrics, and the two others in clinical psychology with experience in neuropsychological evaluation).

Initial translation and pilot administration. In the first step, using a forward-translation strategy (Hambleton & Patsula, 1999), the instructions for all tests were translated to Persian by a qualified bilingual translator who had M.A. degree in English Translation. Next, the translation was examined and corrected in a meeting by a bilingual psychologist and another M.A. translator to ensure the accuracy of the meaning. In the meeting, also, a literally, unrefined translation of the test items was provided. Afterwards, all translated materials were administered in an unstructured, non-standard way to a small sample of 15 individuals representative of the target population. The aim of this initial administration, in which the items were examined instead of collecting data about the subjects (Van de Vijver & Leung, 1997), was to obtain directions for changes needed to reach the adaptation goal. The subjects were asked for their interpretation of instructions, stimuli & items, response alternatives, motivations, and also their alternative suggestions. Consequently, we decided to conduct some local field studies and accordingly change parts of the tests as follows.

Trail Making Test. D-KEFS TMT consists of five conditions: Visual Scanning, Number Sequencing, Letter Sequencing, Number-Letter Switching, and Motor Speed, respectively (for details of the D-KEFS conditions, see the *Examiner Manual*; Delis et al., 2001b). One of the culture-bound features of the test is that it requires knowledge of English alphabet, limiting its use in non-English-speaking populations (Mitrushina, Boone, Razani, & D'Elia, 2005). After several pilot studies, we decided to use the adapted version with Arabic alphabet. Firstly, we used the test with Persian alphabet, but in pilot administrations, our examinees had significant problems. The Persian alphabet is a modified version of the Arabic alphabet, adding four specific letters. In a field study, we designed a condition similar to the D-KEFS Letter Sequencing encompassing the first 16 letters of the Persian alphabet; then, we asked 20 subjects from the population to quickly connect the letters in order, without making mistakes. Interestingly, in addition to a notable delay, the majority of them (14 subjects) skipped at least two letters of three specific Persian letters located on the page. In another field study, in which the impacts of the other variables (e.g., the motor speed) were removed, 15 individuals were asked to write out the letters alphabetically on a blank paper; if they had problem recalling a letter, a page containing scrambled letters could be exposed upon request. All except one of these new 15 subjects missed the specific Persian letters out or used them in an inaccurate alphabetical order. The observed proficiency in the Arabic alphabet is mainly due to the emphasis on the recitation

of this alphabet, as the language of Qur'an, in the introductory religious, Quranic courses that most Baluchs attend. In the final adapted version of the TMT, we used the Arabic script of the letters and numbers.

Verbal Fluency Test. D-KEFS VFT has three conditions: Letter Fluency, Category Fluency, and Category Switching. For a theory-driven adaptation (Malda et al., 2008), in order to extract Persian equivalents for the letters of the first condition (F, A, and S), all Persian letters were ranked based on the frequencies of their respective entry words in two famous Persian dictionaries (as an estimate of the frequency of words that begin with the letters in Persian language); then, from the non-homophonous letters, three with frequencies similar to those for F, A, and S in English (Ohlman, 1959) were selected: *shin*, *kaf*, and *mim*.

Design Fluency Test. D-KEFS DFT is a nonverbal fluency test with three conditions: Filled Dots, Empty Dots Only, and Switching. Given the need to further illustrate the instructions during pilot administrations (as recommended also by Homack et al., 2005), the only modification made to the test was the demonstration of incorrect examples violating the rules, next to the practice boxes of the first condition where the cognitive set of the task is acquired.

Color-Word Interference Test. D-KEFS CWIT consists of four conditions: Color Naming, Word Reading, Inhibition, and Inhibition/Switching. A computerized Persian version of the test (Daryadar, 2014) was utilized in the present study for three reasons. First, given the lengthy time for administering all D-KEFS tests (Baron, 2004) and relatively complex/repetitious test instructions (Homack et al., 2005), as well as the culturally low acceptance of testing among the non-western population (Ardila, 1995), there was a perceived need to reduce cognitive fatigue during pilot administrations, increasing the variety and attractiveness of testing. Second, recording all CWIT variables including error responses (which are informative in TBI assessments) on the paper-pencil version is difficult, requiring the examiner to follow the examinee's progress item by item on both the stimulus booklet and the record form visually and auditorily, thereby increasing the measurement errors. The computerized version records primary and process measures automatically with a precision of hundredths of milliseconds and thereby boosts the reliability of measurement. Third, in addition to a total score for each condition, the computerized version calculates the values individually for all items within CWIT conditions; as a result, it allows using less costly methods of reliability such as internal consistency procedures.

Sorting Test. In D-KEFS ST, two six-card sets are sorted consecutively across two conditions: Free Sorting and Sort Recognition (here, we adapted and used only the standard sets). On each of the card sets, eight different sorting concepts or rules are embedded: three are verbal-semantic (in printed words on the cards) and five are non-verbal (as visual/perceptual features of the cards). Moreover, in the beginning, a practice card set is sorted based on two concepts (names and shape) for training. While generally preserving the original structure and concepts of the test, we made some modifications to the concepts. For the practice set, we replaced the English names with frequently used, indigenous names of boys and girls. For the first testing card set, there were problems in transferring two sorting rules of the stimulus words: "tiger", "car", "airplane", "EAGLE", "BUS", and

"DUCK". First, the sorting rule of "Uppercase Letters vs. Lowercase Letters" does not apply to the Persian script, given that it has no capital letters. We changed this *perceptual* (non-verbal) rule to "Bold letters vs. non-Bold letters", boldfacing the uppercase words in the Persian version. Second, yielding syllabically various words, literal translations of the stimulus words could not be sorted based on the rule of "One-Syllable Words vs. Two-Syllable Words." For this rule, we changed two words on the cards into new words within the same semantic category and at the same level as the originals; thereby, we assembled a new rule: "Two-Syllable Words vs. Four-Syllable Words." In order to preserve the original arrangement of the concepts and prevent repeating a target array (sort), the new rule was mounted on the array of one of the non-target even sorts. Words in the final version were selected based on iterative pilot administrations and field surveys for familiarity/recognizability.

Twenty Questions Test. The 20-Question is a well-known game and radio quiz program in Iran. We only translated the instructions of D-KEFS TQT to Persian. Our participants were well-familiar with the stimulus pictures.

Word Context Test. D-KEFS WCT consists of 10 items, on each of which the examinee attempts to guess the meaning of a made-up word based on its use in five successive clue sentences. The first step in development of the Persian WCT for our target population was to include *pseudowords* phonologically and graphemically dissimilar to (not reminder of) a real word in both Baluchi and Persian languages, as well as inflectable in the Persian syntax. Therefore, new made-up words were substituted for seven of the problematic original pseudowords, with the help of a native linguist and 15 bilingual Baluchs from various counties of the province. For the *mystery* words (the target meanings) and their respective clue sentences, various changes were made during the iterative adaptation process. The use of literal translations for items of the mystery verbs (e.g., *make*, *eat*) was problematic, given that the corresponding made-up words had to be changed graphemically when placed and inflected in Persian clue sentences. Thus, instead, we used the Persian verbs that have an invariable/uninflected meaningful part and also could have the same function as the English verbs; then, we created new clue sentences for these verbs, in which the invariable part was replaced with the made-up word. Some other changes were made. For example, there is no word in Persian with the same meaning as *voice* (one of the mystery meanings) in English. Instead, we used the Persian word *nâm* or *esm* (name) and created new corresponding sentences for it. All clue sentences in the test, especially those with English idioms, were adapted to lingual, social, and cultural characteristics of the population.

Tower Test. D-KEFS TT is a non-verbal test. For this test, only the test instructions were required to be translated (as concluded from pilot studies).

Proverb Test. In D-KEFS PT, eight common and uncommon proverbs are presented to the examinee in two formats: Free Inquiry and Multiple Choice. Two of the five common sayings (the *chickens* and the *cooks*) have appropriate Persian equivalents; for the three others, we had to choose the Persian proverbs common in the region. For this purpose, in a field survey, we asked 125 native individuals to select common proverbs in the region from a list of 100 common Persian sayings.

Accordingly, we chose 12 frequently selected proverbs, which also met the psychometric requirements of D-KEFS (with two extractable concrete and abstract meanings, multiple separable elements, etc.); from among them, three proverbs were selected as a result of pilot administrations of the test. For the uncommon proverbs of the test, the commonness was not the prerequisite; thus, we translated three foreign proverbs (one from the proverbs of the test, one from Denmark, and the other from Germany), the conceptual elements of which were familiar to the population under study.

BADS. The Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) is a test battery designed to predict everyday dysfunctions that arise from the dysexecutive syndrome, especially among brain damaged patients. BADS possesses very high ecological validity (Crawford & Henry, 2005); In order to address the problem of low ecological validity of traditional executive tests, it assesses executive skills and demands similar to everyday activities while maintaining a structured format. The battery consists of six tests: Rule Shift Cards (RS), Action Program (AP), Key Search (KS), Temporal Judgment (TJ), Zoo Map (ZM), and Modified Six Elements Task (6E). Satisfactory psychometric properties of BADS have been established in various studies (e.g., Espinosa et al., 2009; Gouveia, Brucki, Malheiros, & Bueno, 2007; Norris & Tate, 2000; Wilson et al., 1996, 1998). Given that there were no other executive tests previously translated or developed for the population of the current study (to be used for convergent validation of the adapted tests), we selected BADS and translated the test instructions using the same forward strategy conducted for the D-KEFS instructions (Hambleton & Patsula, 1999). Then, concluded from a non-standard administration of the battery to 15 native individuals, only the items of TJ, a culturally biased test (Chamberlain, 2003) required to be adapted, replacing the items with culturally familiar events. In the final step, we conducted a pilot standard administration of the tests to 20 individuals representative of the population and accordingly, we removed two tests of the BADS: TJ and AP. TJ had no relation with the other BADS tests and the BADS total score. Similarly, previous research have demonstrated poor psychometric properties for this test (e.g., Bennett, Ong, & Ponsford, 2005; Gillespie, Evans, Gardener, & Bowen, 2002; Norris & Tate, 2000). The other inconsistent test was AP, not related to the other tests, with severe negative skewness and no variance in its scores. Consequently, we included only four tests of BADS (i.e., RS, KS, ZM, and 6E) in the final protocol of this study. BADS also has a Dysexecutive Questionnaire (DEX) that was not utilized in this study.

Phase 2 - A posteriori procedures: statistical analysis of reliability and validity

Participants

Using the purposive-homogenous sampling, 83 individuals who had the aforementioned characteristics of the population under study and endorsed no history of the exclusion medical and psychiatric symptoms/conditions (e.g., head injuries, substance abuse, colour-blindness; see Delis et al., 2001c) were selected in

public gathering places of five cities in the region. In order to further screen for normality, a validated Persian version (Seyedian et al., 2007) of the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) was administered to all the individuals and those with total scores of 24 or less were excluded. Accordingly, 75 subjects were included in final administration. During the administration, two of the subjects had unusual testing behaviors; subsequent investigations revealed psychiatric disorders in both (ADHD and PTSD, previously undiagnosed due to lack of awareness and facilities). The data for both the subjects were excluded from reliability and validity analyses. Therefore, the analyses were conducted on the data for 73 healthy, young Baluch men aged 16 to 40 years (*M*= 24.34, *SD*= 6.53) with 3 to 12 years of education (*M*= 9.33, *SD*= 2.47). The study was approved by the institutional ethics committee and conducted in compliance with the Helsinki declaration; all the participants gave informed consent prior to participation.

Measures

The adapted materials for eight D-KEFS tests described above, along with the computerized Persian D-KEFS CWIT (Daryadar, 2014) were administered. The computerized CWIT presents the same practice and testing items as in the paper-pencil version, individually in the center of the computer screen. The examinee is instructed to press the corresponding key for each color as quickly as s/he can without making mistakes. The software calculates reaction times (RTs) and errors for the testing items. Almost all participants in the current study had no prior computer experience. Moreover, as stated previously, we used the four tests from BADS (the RS, KS, ZM, and 6E) with Persian instructions. For each of the BADS tests, a profile score was calculated (0 to 4), and then we added up the profile scores of the four tests to obtain a BADS total profile score (BADS4t). For the 73 participants, the Cronbach's alpha between the four BADS tests was .70 and the mean inter-test correlation was .39.

Administration procedure

The participants were individually tested in a quiet, distraction-free room over a session lasting between 2-3 hours with a 20-minute break after the first five tests. As indicated previously, one of the main troubles we had in the data collection process was low cooperativeness and acceptance of testing, a common characteristic of low socioeconomic populations. Given this fact and the lengthy time of administration in Phase 2, besides using the Computerized CWIT and the break, we made two of the less psychometrically established D-KEFS tests optional (WCT and PT; these tests were also not included in our assessment of TBI patients, Ghawami et al., 2016) and placed them in the end of the protocol, in order to alleviate the impact of cognitive fatigue and interest/effort reduction. Therefore, the tests were administered based on their standardized order with an exception: first, D-KEFS TMT, VFT, DFT, CWIT, ST, TQT, and TT were consecutively administered; then, BADS RS, KS, ZM, and 6E tests; and lastly, the optional tests,

D-KEFS WCT and PT. Of our participants, 28 individuals did not respond to the two optional tests. Thus, those analyses that contained the optional tests were based on data for 45 participants. There were no significant differences between the two groups of the participants on the other D-KEFS and BADS tests. Twenty-four subjects were retested within 12 to 30 days (*M*= 18 days) on the tests for which the reliability cannot be estimated via a single testing session.

Data analysis

The scores were normally distributed for the majority of variables. Using analyses similar to those in the original standardizations study (Delis et al., 2001c), various correlation coefficients were calculated. In the reliability analysis, internal consistency and/or stability methodologies were utilized depending on the nature and procedures of the D-KEFS tests: For VFT, ST, TQT, TT, WCT, and PT, split-half internal consistencies were computed and corrected by the Spearman-Brown formula; for the CWIT conditions, internal consistencies were calculated using the Cronbach's alpha coefficient; afterwards, test-retest reliabilities were derived for TQT (using the alternate form), TMT, DFT, and TT. The size of reliability coefficients was interpreted according to the guideline by Strauss et al. (2006): *low* (< .59), *marginal* (.60 to .69), *adequate* (.70 to .79), *high* (.80 to .89), and *very high* (.90+). In the validity analysis, intercorrelations of measures within individual D-KEFS tests and correlations between measures of different D-KEFS tests were computed; also, correlations of D-KEFS tests with the BADS scores were calculated. The validity values were interpreted according to the following general labels (Cohen, 1988): *weak* ($r < .30$), *moderate* ($.30 \leq r < .50$), and *strong* ($.50 \leq r$).

Results

There were no effects for the key demographic variables, such as age and education. This is not surprising given the demographic homogeneity of the current sample. Descriptive statistics for selected D-KEFS and BADS variables, including means and standard deviations, are displayed in Table 1.

Table 1
Descriptive statistics for D-KEFS and BADS raw scores

Measure	*M*	*SD*
D-KEFS		
Trail Making 1: Visual Scanning time (s)	24.56	5.53
Trail Making 2: Number Sequencing time (s)	37.99	13.61
Trail Making 3: Letter Sequencing time (s)	67.26	35.59
Trail Making 4: Number-Letter Switching time (s)	116.49	54.31
Trail Making 5: Motor Speed time (s)	34.89	13.75
Verbal Fluency 1: Letter Fluency, Total Correct	26.53	9.61
Verbal Fluency 2: Category Fluency, Total Correct	45.15	9.57
Verbal Fluency 3: Category Switching, Total Correct Responses	15.11	3.16
Verbal Fluency 3: Category Switching, Total Switching Accuracy	14.33	3.31

Measure	*M*	*SD*
Design Fluency 1: Filled Dots, Total Correct	9.27	3.02
Design Fluency 2: Empty Dots Only, Total Correct	10.19	3.58
Design Fluency 3: Switching, Total Correct	7.85	2.30
Design Fluency: Total Correct	27.32	7.80
Color-Word Interference 1: Color Naming, Mean Reaction Time (ms)	758.65	243.96
Color-Word Interference 2: Word Reading, Mean Reaction Time (ms)	740.38	193.88
Color-Word Interference 3: Inhibition, Mean Reaction Time (ms)	1019.09	364.66
Color-Word Interference 4: Inhibition/Switching, Mean Reaction Time (ms)	1784.10	519.28
Sorting 1: Free Sorting, Confirmed Correct Sorts	8.29	2.42
Sorting 1: Free Sorting, Description Score	30.36	9.45
Sorting 2: Sort Recognition, Description Score	30.30	8.82
Twenty Questions: Initial Abstraction Score	29.01	11.11
Twenty Questions: Total Questions Asked	29.86	10.57
Twenty Questions: Total Weighted Achievement Score	14.32	3.72
Word Context: Total Consecutively Correct	22.24	7.14
Word Context: Total Repeated Incorrect Responses	6.40	4.38
Tower: Total Achievement Score	16.11	3.90
Tower: Total Rule Violations	2.75	3.42
Proverbs 1: Free Inquiry, Total Achievement Score	23.27	5.54
Proverbs 2: Multiple Choice, Total Achievement Score	28.84	4.72
BADS		
Rule Shift Cards, Profile Score	3.26	0.80
Key Search, Profile Score	2.62	1.24
Zoo Map, Profile Score	2.78	1.11
Modified Six Elements, Profile Score	3.41	0.77
BADS4t (BADS Total Profile Score–4subtest)	12.07	2.89

Note: D-KFES= Delis-Kaplan Executive Function System; BADS= Behavioural Assessment of the Dysexecutive Syndrome; RT= Reaction Time; s= seconds; ms= milliseconds.

Reliability

In the internal consistency analysis, the same split-half methodologies as in the original standardization (Delis et al., 2001c) were conducted for most of the tests (comparing performance on the even-odd time intervals for VFT; correlating performance on the two card sets for ST; and treating even and odd items of TQT, TT, WCT, and PT as equivalent half tests). For CWIT, we used the Cronbach's alpha coefficient, given that the software individually calculates reaction times for all the 50 items. In tests for which internal consistency cannot be assessed (TMT and DFT) or may be attenuated (TQT and TT) due to item interdependence (Delis et al., 2001c), test-retest reliabilities were calculated. The internal consistency and test-retest reliability values for the primary scores of the D-KEFS tests are presented in Table 2. Most of the reliabilities were in the adequate to very high range. The highest internal consistencies were found for the CWIT conditions (.96-.97), followed by VFT Letter Fluency (.87), TQT Initial Abstraction (.85), and the Sorting

scores (.81-.84); and the lowest were for the achievement scores of TT (.44) and TQT (.41). There were performance improvements on all the primary scores from the first to the second testing. The highest stability coefficients were found for TQT Initial Abstraction (.92), TMT Number-Letter Switching (.90), and DFT Filled Dots (.88); and the lowest stability was found for DFT Switching (.48). Indicating the effect of item interdependence, test-retest reliabilities were clearly higher than internal consistencies for tests for which both the methods were utilized (TQT and TT), although their stability values were in the marginal range and still were not adequate for the achievement scores of these test (.62 and .61, respectively).

Table 2
Reliability values for primary D-KEFS measures

Measure	Internal consistency[a]	Test	Retest	r_{12}
		M (*SD*)	*M* (*SD*)	
TMT: Visual Scanning (time)		25.50 (6.04)	23.75 (4.87)	.81
TMT: Number Sequencing (time)		37.58 (9.69)	35.04 (5.62)	.59
TMT: Letter Sequencing (time)		67.17 (34.91)	48.13 (21.96)	.86
TMT: Number-Letter Switching (time)		109.42 (48.90)	85.25 (31.76)	.90
TMT: Motor Speed (time)		36.04 (11.10)	32.00 (7.16)	.71
VFT: Letter Fluency, Total Correct	.87			
VFT: Category Fluency, Total Correct	.58			
VFT: Category Switching, Correct Responses	.60			
VFT: Category Switching, Switching Accuracy	.69			
DFT: Filled Dots, Total Correct		8.67 (2.28)	10.83 (2.48)	.88
DFT: Empty Dots Only, Total Correct		9.17 (2.75)	10.79 (2.96)	.86
DFT: Switching, Total Correct		7.88 (2.05)	8.42 (1.86)	.48
CWIT: Color Naming (time)	.97			
CWIT: Word Reading (time)	.97			
CWIT: Inhibition (time)	.97			
CWIT: Inhibition/Switching (time)	.96			
ST: Free Sorting, Confirmed Correct Sorts	.82			
ST: Free Sorting, Description Score	.84			
ST: Sort Recognition, Description Score	.81			
TQT: Initial Abstraction Score	.85	32.17 (11.98)	32.92 (11.21)	.92
TQT: Total Questions Asked	.74	28.29 (4.83)	28.08 (5.64)	.78
TQT: Total Weighted Achievement Score	.41	14.92 (2.41)	15.17 (2.90)	.62
WCT: Total Consecutively Correct	.78			
TT: Total Achievement Score	.44	16.46 (3.54)	17.96 (3.32)	.61
PT: Free Inquiry, Total Achievement Score	.68			
PT: Multiple Choice, Total Achievement Score	.76			

Notes: D-KFES= Delis-Kaplan Executive Function System; TMT= Trail Making Test; VFT= Verbal Fluency Test; DFT= Design Fluency Test; CWIT= Color-Word Interference Test; ST= Sorting Test; TQT= Twenty Questions Test; WCT= Word Context Test; TT= Tower Test; PT= Proverbs Test. [a]The values for the CWIT are Cronbach's alpha coefficients; the other values are split-half coefficients.

Validity

INTERCORRELATIONS OF MEASURES WITHIN INDIVIDUAL D-KEFS TESTS. There were significant positive correlations between the Trail Making primary measures (Table 3). The correlations were generally in the moderate to strong range. The three sequencing conditions had strong positive relationships with each other (.54-.74, $p < .001$). The more basic conditions (Visual Scanning and Motor Speed) had relatively lower correlations with the sequencing conditions (.28-.57).

Table 3
Intercorrelations of D-KEFS Trail Making Test (completion times)

Measure	Number Sequencing	Letter Sequencing	Number-Letter Switching	Motor Speed
Visual Scanning	.55**	.44**	.37**	.52**
Number Sequencing		.61**	.54**	.57**
Letter Sequencing			.74**	.39**
Number-Letter Switching				.28*

Note: *$p < .05$, two-tailed.**$p < .01$, two-tailed.

Strong positive associations were found between the three Verbal Fluency conditions (Correct Responses). The highest association was among Category Fluency and Category Switching ($r = .63$, $p < .001$). The two category conditions were comparably associated with Letter Fluency ($r = .53$ and .55, respectively, $p < .001$). In addition, a very strong correlation was found between the two primary measures of the Category Switching condition ($r = .97$, $p < .001$).

The relations among Design Fluency Total Correct scores were significant and generally in the strong range. The correlation between the Filled Dots and Empty Dots Only conditions ($r = .81$, $p < .001$) was higher than their respective correlations with the Switching condition ($r = .47$ and .58, respectively, $p < .001$). The test's composite score, Design Fluency Total Correct was more associated with the two no-switching conditions ($r = .90$ and .94, respectively, $p < .001$) than the switching condition ($r = .74$, $p < .001$).

There were strong relations between all the conditions of the computerized Color-Word Test (Table 4). The highest correlation was between the two baseline conditions (Color Naming and Word Reading). The two higher-level conditions had a strong positive correlation with each other ($r = .65$, $p < .001$).

Very strong relationships were found between the Sorting primary measures. Free Sorting Confirmed Correct Sorts was positively and highly correlated with both Free Sorting Description Score ($r = .96$, $p < .001$) and Sort Recognition Description Score ($r = .84$, $p < .001$). In addition, the two description scores were strongly related to each other ($r = .87$, $p < .001$).

Table 4
Intercorrelations of D-KEFS CWIT

Measure	Word Reading	Inhibition	Inhibition/Switching
Color Naming (mRT)	.90**	.74**	.69**
Word Reading (mRT)		.76**	.65**
Inhibition (mRT)			.65**

Notes: CWIT= Color-Word Interference Test; mRT= mean Reaction Time. **$p < .01$, two-tailed.

There were moderate to strong relations among the Twenty Questions primary measures, suggesting that good performances on the measures are associated with each other. TQT Initial Abstraction Score was significantly and moderately correlated with Total Questions Asked (r= -.44, p< .001) and Total Weighted Achievement Score (r= .38, p= .001). A strong negative correlation was found between Total Weighted Achievement Score and Total Questions Asked (r= -.84, p< .001).

For Word Context, we calculated the correlation between Total Consecutively Correct (the test's primary achievement measure) and Total Repeated Incorrect Responses. A strong negative correlation was obtained between the measures (r= -.56, p< .001), indicating that higher achievement score is associated with lower number of repeated incorrect responses.

For the tower, we computed the correlation between Total Achievement Score and Total Rule Violations. A strong negative correlation was found (r= -.61, p< .001), suggesting that better overall performance on the test is associated with fewer rule violations.

There was a positive and significant moderate correlation between the Accuracy Only and Abstraction Only scores of Proverbs Test Free Inquiry (r= .39, p= .008). Total Achievement Score had strong correlations with Accuracy Only (r= .80, p< .001) and Abstraction Only (r= .83, p< .001) measures.

CORRELATIONS BETWEEN MEASURES OF DIFFERENT D-KEFS TESTS. Correlations among various primary executive measures from different D-KEFS tests are displayed in Table 5. The results demonstrate that executive scores within the same test are related more strongly than executive scores across tests. Similar to those of Delis et al. (2001c), the correlations between tests were generally weak to moderate, with the majority of them in the weak range; only four between-test correlations were in the strong range.

RELATIONS OF D-KEFS TESTS WITH THE BADS. Table 6 presents correlations between the D-KEFS tests and the selected BADS measures. Most of the relations were statistically significant and in the moderate range. The negative values indicate the association between better performances. The most powerful correlations of D-KEFS tests were with BADS4t, where all D-KEFS measures had significant, moderate to strong correlations with BADS4t. The highest correlations of BADS4t were with TT Total Achievement Score (r= .62, p< .001), ST Sort Recognition Description Score (r= .57, p< .001), and TMT Number-Letter Switching (r= -.51, p< .001). Excluding TQT Weighted Achievement Score, all the D-KEFS measures were significantly correlated with 6E. Moreover, with the exception of CWIT inhibition and WCT, all D-KEFS measures had significant correlations with ZM. A pattern of correlations was revealed between the VFT and BADS subtests: Whereas all the VFT measures were significantly correlated with ZM and 6E, none of them were correlated with the RS and KS. The same pattern was revealed for the PT achievement score.

Table 5

Correlations between D-KEFS Tests (achievement scores)

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Trail Making, Number-Letter Switching (time)	-.18	-.23*	-.20	-.22	-.51**	.46**	.61**	-.41**	-.45**	-.46**	-.40**	-.20	-.54**	-.25
2. Verbal Fluency, Letter Fluency, Total Correct	1	.53**	.55**	.56**	.29*	-.21	-.26*	.37**	.36**	.30*	.21	.32*	.22	.31*
3. Verbal Fluency, Category Fluency, Total Correct		1	.62**	.63**	.30**	-.10	-.20	.19	.18	.15	.14	.06	.18	.53**
4. Verbal Fluency, Category Switching, Correct Responses			1	.97**	.38**	-.26*	-.16	.19	.19	.14	.06	.14	.20	.17
5. Verbal Fluency, Category Switching, Switching Accuracy				1	.38**	-.20	-.15	.19	.18	.16	.12	.14	.21	.21
6. Design Fluency, Total Correct					1	-.41**	-.45**	.44**	.46**	.42**	.29*	.25	.42**	.15
7. CWIT, Inhibition, Mean Reaction Time						1	.65**	-.22	-.23	-.27*	-.26*	-.02	-.32**	.00
8. CWIT. Inhibition/Switching, Mean Reaction Time							1	-.29*	-.31*	-.33**	-.45**	-.17	-.37**	-.27
9. Sorting, Free Sorting, Total Confirmed Correct Sorts								1	.96**	.84**	.47**	.41**	.35**	.20
10. Sorting, Free Sorting, Total Description Score									1	.87**	.49**	.39*	.35**	.19
11. Sorting, Sort Recognition, Total Description Score										1	.49**	.48**	.44**	.10
12. Twenty Questions, Total Weighted Achievement Score											1	.28	.32**	.12
13. Word Context, Total Consecutively Correct Score												1	.11	.10
14. Tower Test, Total Achievement Score													1	.00
15. Proverbs, Free Inquiry, Total Achievement Score														1

Notes: For the Trail Making and CWIT lower scores indicate better performance (the reason for the negative values). CWIT= Color-Word Interference Test. *p< .05, two-taile;.**p< .01, two-tailed.

Table 6
Correlations between D-KEFS Achievement Scores and BADS Scores

	BADS				
	RS	KS	ZM	6E	BADS4t
Trail Making Test (TMT)					
Number-Letter Switching (time)	-.44**	-.39**	-.35**	-.34**	-.51**
Verbal Fluency Test (VFT)					
Letter Fluency, Total Correct	.20	.19	.31*	.42**	.36**
Category Fluency, Total Correct	.10	.21	.35**	.25*	.32**
Category Switching, Correct Responses	.15	.17	.26*	.32**	.30*
Category Switching, Switching Accuracy	.19	.13	.28*	.33**	.31*
Design Fluency Test (DFT)					
Filled Dots, Total Correct	.36**	.24*	.36**	.38**	.44**
Empty Dots Only, Total Correct	.35**	.22	.28*	.36**	.39**
Switching, Total Correct	.34**	.31**	.33**	.31*	.43**
Color-Word Interference Test (CWIT)					
Inhibition, mean Reaction Time	-.11	-.16	-.21	-.47**	-.31**
Inhibition/Switching, mean Reaction Time	-.03	-.36**	-.34**	-.41**	-.42**
Sorting Test (ST)					
Free Sorting, Total Confirmed Correct Sorts	.34**	.26*	.36**	.33**	.43**
Free Sorting, Total Description Score	.38**	.27*	.36**	.36**	.46**
Sort Recognition, Total Description Score	.44**	.35**	.45**	.47**	.57**
Twenty Questions Test (TQT)					
Initial Abstraction Score	.15	.27*	.44**	.24*	.39**
Total Questions Asked	-.21	-.38**	-.51**	-.26*	-.49**
Total Weighted Achievement Score	.13	.29*	.46**	.22	.39**
Word Context Test (WCT)[a]					
Total Consecutively Correct Score	.17	.27	.25	.36*	.37*
Tower Test (TT)					
Total Achievement Score	.42**	.47**	.50**	.43**	.62**
Proverbs Test (PT)[a]					
Free Inquiry, Total Achievement Score	.21	.17	.41**	.30*	.39**

Notes: For the Trail Making, Color-Word, and Questions Asked on the Twenty Questions, lower scores indicate better performance (the reason for the negative values). BADS= Behavioural Assessment of the Dysexecutive Syndrome; RS= Rule Shift Cards; KS= Key Search; ZM= Zoo Map; 6E= Modified Six Elements; BADS4t= BADS Total Profile Score–calculated from the four tests. [a]n= 45; *p< .05; two-tailed; **p< .01, two-tailed.

Discussion

As a response to the need for adapting well-established western neuropsychological tests for assessments of TBI patients in Iran, particularly in the south-eastern region, the current study was aimed mainly to develop and preliminarily validate a Persian version of the D-KEFS tests for assessments of TBI patients in Zahedan, Iran. In the judgmental phase (Phase 1), we attempted to provide cross-culturally appropriate materials, conducting several non-standard pilot administrations and local field studies. Consequently, in addition to the forward translation of test instructions, we made several changes to the test items. The content adaptation efforts were under advisement, supervision, and content judgments of a mixed committee of experts. For the statistical phase (Phase 2), we preliminarily validated the provided materials, examining their reliability and validity on a normal sample with demographic traits of the TBI patients.

In the reliability analysis, except for CWIT, the same internal consistency and test-retest methodologies as in the original standardization study of D-KEFS (Delis et al., 2001c) were conducted. The results demonstrated adequate to high split-half and stability coefficients for most of the measures. Although generally with the same pattern, the coefficients were mostly higher than their counterparts for corresponding age groups in the US standardization study.

Schmidt (2003) criticized the reported reliabilities in the D-KEFS manual, pointing out that only 17% of the values were above a .80 value. Delis et al. (2004) argued that a reason for the lower reliability values could be the greater performance variability in the complex executive tests which tap a wider spectrum of cognitive processes. Given the complex, multidimensional nature of executive functions, this neuropsychological domain is difficult to assess reliably (Strauss et al., 2006). Executive function tests often show low internal and stability reliabilities (Alvarez & Emory, 2006; Chung, Weyandt, & Swentosky, 2014; Miyake et al., 2000); this is the case particularly for test-retest reliabilities in view of the crucial demand of executive tasks on the novelty/effortful processing rather than the practiced/automatic processing (Phillips, 1997; Rabbitt, 1997). Another reason for the low reliabilities in past studies may be that examinees use different strategies on different occasions, even within the same session, when responding to the executive tests (Miyake et al., 2000).

Some factors may have influenced the reliabilities in the current study. In order to limit the effects of inter-examiner variations, the administration and scoring of performances for all the examinees were only done by one trained examiner (the first author). Moreover, despite the demographic homogeneity of the current sample, noticeable variances were found for the target executive variables, allowing for the relatively higher reliability coefficients. On the other hand, it should be noted that in contrast to what has been taken for granted by most measurement textbooks and discussions in the literature, higher reliability is not solely associated with higher true score variability and corresponding sample heterogeneity, but also with error variance and consistency of within-person factors (especially in case of executive functions); specifically, it is a function of lower ratio of error variance to true variance (Feldt & Qualls, 1999). In assessments

of complex executive functions, provided that there is a reasonable level of true score variability (e.g., in clinical or young samples), sometimes sample homogeneity in terms of noisy variables can attenuate unwanted variability and thereby increase reliability. For example, for the current sample, adding individuals from the middle socioeconomic class could result in various levels of benefit from practice, and thereby decrease the test-retest correlations.

For reliability of the CWIT, given the individual reaction times for all items, we used the Cronbach's alpha method. Very high alpha coefficients were obtained for all the conditions (.96-.97). The computerized CWIT records measures automatically and precisely, and hence boosts the reliability of measurement. In addition to their economic advantages, the use of computerized tests decreases the possibility of administration and scoring errors (Mataix-Cols & Bartes-Faz, 2002; Parsey & Schmitter-Edgecombe, 2013). Therefore, the use of computerized tests can be seen among remedies for the low reliability of executive tests.

In the validity analysis, intercorrelations within the tests were of patterns relatively similar to, but mostly higher than, their counterparts for corresponding age groups in the original standardization study (Delis et al., 2001c). As expected, strong significant correlations were found between the majorities of the executive scores within the tests, indicating that intratest measures are related to the same construct. On the other hand, consistent with the findings of Delis et al (2001c), correlations between executive scores from various D-KEFS tests were generally weak. Delis and colleagues suggested that these results "indicate that instruments are not interchangeable and measure unique aspects of executive functioning with some overlap of variance" (p. 82). This is also consistent with the body of literature, where correlations between putative executive tasks usually have been low, often non-significant, and hence, construct validity of executive functioning as a monolithic construct has been in question (Alvarez & Emory, 2006; Miyake et al., 2000; Rabbitt, 1997; Salthouse, Atkinson, & Berish, 2003).

There may be several reasons for the low correlations in the current and previous studies. The low reliabilities of executive tests may attenuate their intercorrelations (Miyake et al., 2000): a variable with low reliability has a small proportion of its variance available for associations with other variables (Salthouse et al., 2003); the upper limit for correlation between two variables is equal to the geometric mean of their reliabilities (Ferguson & Takane, 1989). The weak correlations between the D-KEFS tests in the current study are not solely attributable to reliability, given the higher reliabilities for most of the executive measures. For example, while VFT Letter Fluency and CWIT Inhibition primary scores had high reliabilities (.87 and .97, respectively; with the maximum possible correlation of .92 determined by the reliabilities), the correlation between them was weak and non-significant (-.21). Another reason for the low correlations between executive tasks in the literature may be a widely discussed problem, "task impurity" (Burgess, 1997; Hughes & Graham, 2002; Phillips, 1997). That is to say, in view of the fact that complex executive tasks involve multiple executive and non-executive processes, the relations are not simply between purely measured executive functions. Generally, the pattern of within- and between-test correlations for the D-KEFS measures in the current study is more consistent with a

connectionist explanation (McCloskey, Perkins, & Van Divner, 2009; Rabbitt, 1997). In other words, it is more plausible to envisage the managerial role of executive functions within the global network of cognition, an information-processing network including a wide range of domain-/task-specific subsystems transversely and hierarchically interconnected with each other. In this view, it is also legitimate to envisage substantial theoretical, functional, and neuroanatomical overlaps between executive functions and general cognitive ability factors, such as psychometric *g* (e.g., Barbey, Colom, & Grafman, 2013; Barbey et al., 2012; Duncan et al., 2000; Keifer & Tranel, 2013). Given these interconnections and the nature of executive functioning, involving simultaneous co-ordination of various cognitive processes, the classical strategy of isolating a single executive index, i.e., experimental control of demand variables seems inappropriate (Hughes & Graham, 2002; Rabbitt, 1997).

In the validity analysis with BADS, most of the correlations between D-KEFS and BADS measures were statistically significant and in the moderate range. Correlations of D-KEFS with BADS Total Profile Score (BADS4t) were higher than those with the BADS subtests: All the D-KEFS measures had significant, moderate to strong correlations with BADS4t. Considering the aforementioned inherent problems in the construct validation of executive tests, these results can be seen as an acceptable evidence of convergent validity for the D-KEFS tests. The relations can also be viewed as an indirect estimation of ecological validity for the D-KEFS tests, given that BADS was designed primarily to predict everyday dysfunctions and the tests are similar to real-life activities; for example, losing something, like the keys in KS, is a fairly commonplace event in the life of most people, particularly TBI patients (Wilson et al., 1996, 1998). One possible reason for the relatively lower correlations of D-KEFS with the BADS subtests compared with BADS4t, may be the limited range in possible profile scores of the subtests (0-4). The non-uniform pattern of correlations between the D-KEFS executive measures and the BADS subtests supports the above discussed, non-unitary nature for executive functions.

Delis, Jacobson, Bondi, Hamiltoon, and Salmon (2003) showed serious shortcomings of factor analytic methods in construct validations of process-oriented cognitive tests with normal or mixed-clinical samples. This was the main reason for not using factor analyses and across-test composite indices in the development of D-KEFS (Delis et al., 2004). In addition to the shortcomings, for the current preliminary study, although the sample size was fair for such a demographically homogeneous sample and met the adequate value of 50 subjects for a normative sample (Mitrushina et al., 2005), it was small for exploratory or confirmatory factor analyses.

Puente and Perez-Garcia (2000) pointed out that it may be difficult to obtain permission from Asians or Hispanics to examine their minds. Indeed, given the low acceptance of testing in non-western countries and cultures (Ardila, 1995), particularly in the current low-socioeconomic population, sampling among the lower-literate men, who are mostly school dropouts, for a lengthy cognitive testing is laborious. It was very difficult for us to persuade a Baluch full-time laborer to give some hours of his time responding to the inherently "unpleasant", effortful, and challenging executive tests (Phillips, 1997). For data collection, concluding

from pilot administrations, we took some precautions to alleviate the impact of cognitive fatigue and interest/effort reduction during testing: using a computerized test after administering the first three tests in order to reduce the monotony, placing a break time in the middle of administration, and making the last two tests (WCT and PT) optional. Only 45 participants responded to the optional tests. Therefore, cautions should be taken when using or interpreting our results for these two tests.

Another sample-related limitation of the present study involves the generalizability of the findings to other populations, given the homogeneity of the current sample.

In conclusion, generally, the result of this preliminary validation study showed relatively high reliability and acceptable convergent validity for this Persian version of D-KEFS. The patterns of the correlations in the validity analysis were consistent with the body of literature, demonstrating a multidimensional, non-unitary nature for executive functioning. Further validity evidence, including clinical utility of the tests with TBI patients, was obtained in another study, where TBI patients with focal frontal lesions in Khatam-ol-anbia Hospital showed substantial impairments compared with the current normal sample on all the primary executive measures (Ghawami et al., 2016).

References

Altemeier, L., Jones, J., Abbott, R. D., & Berninger, V. W. (2006). Executive functions in becoming writing readers and reading writers: Note taking and report writing in third and fifth graders. *Developmental Neuropsychology*, *29*, 161-173.

Alvarez, J. A., & Emory, E. (2006). Executive functions and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, *16*, 17-42.

Ansari-Moghaddam, A., Martiniuk, A. L. C., Mohammadi, M., Rad, M., Sargazi, F., Sheykhzadeh, K., Jelodarzadeh, S., & Karimzadeh, F. (2012). The pattern of injury and poisoning in South East Iran. *BMC International Health and Human Rights, 12*, 17.

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 143-150.

Ardila, A. (2007). The impact of culture on neuropsychological test performance. In B. P. Uzzell, M. O. Pontón, & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology* (pp. 23-44). Mahwah, NJ: Erlbaum.

Barbey A. K., Colom R., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*, *51*, 1361-1369.

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain, 135*(4), 1154-1164. doi:10.1093/brain/aws021

Baron, I. S. (2004). Delis-Kaplan Executive Function System [Test review]. *Child Neuropsychology*, *10*, 147-152.

Bennett, P. C., Ong, B., & Ponsford, J. (2005). Measuring executive dysfunction in an acute rehabilitation setting: Using the dysexecutive questionnaire (DEX). *Journal of the International Neuropsychological Society*, *11*, 376-385.

Bhalla, K., Sharaz, S., Abraham, J. P., Bartels, D., & Yeh, P. H. (2011). *Road injuries in 18 countries.* Boston, MA: Harvard University.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81-116). Hove: Psychology Press.

Chamberlain, E. (2003). Behavioural Assessment of the Dysexecutive Syndrome (BADS) [Test review]. *Journal of Occupational Psychology, Employment and Disability*, *5*, 33-37.

Chardoli, M., & Rahimi-Movaghar, V. (2006). Analysis of trauma outcome at a university hospital in Zahedan, Iran using the TRISS method. *East African Medical Journal*, *83*, 440-442.

Chung, H. J., Weyandt, L. L., & Swentosky, A. (2014). The physiology of executive functioning. In S. Goldstein, & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 13-27). New York, NY: Springer

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

Crawford, J. R., & Henry, J. D. (2005). Assessment of executive dysfunction. In P. W. Halligan, & D. T. Wade (Eds.), *The effectiveness of rehabilitation for cognitive deficits* (pp. 233-245). London: Oxford University Press.

Crawford, J. R., Sutherland, D., & Garthwaite, P. H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society*, *14*, 1069-1073.

Daryadar, M. (2014). *Barresi negareshe fard nesbat be sedaye zabt-shodeye khod va taa'sire sedaye zabt-shodeye afrad bar yadgiriye shenidari-kalami va tureshe tavajoh* [Investigating individual's attitude toward own recorded voice and the effect of listening to recordings of one's voice on attentional bias and auditory verbal learning] (Unpublished master's thesis). University of Sistan and Baluchestan, Zahedan, Iran.

Davis, A. S., Peirson, E. E., & Finch, W. H. (2011). A canonical correlation analysis of intelligence and executive functioning. *Applied Neuropsychology*, *18*, 61-68.

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using shared variance techniques with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*, 936-946.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001a). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001b). *Delis-Kaplan Executive Function System (D-KEFS): Examiner's manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001c). *Delis-Kaplan Executive Function System (D-KEFS): Technical manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society*, *10*, 301-303.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N., & Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*, 457-460.

Espinosa, A., Alegret, M., Boada, M., Vinyes, G., Valero, S., Martínez-Lage, P., Peña-Casanova, J., Becker, J. T., Wilson, B. A., & Tarraga L. (2009). Ecological assessment of executive functions in mild cognitive impairment and mild Alzheimer's disease. *Journal of the International Neuropsychological Society*, *15*, 751-757.

Feldt, L. S., & Qualls, A. L. (1999). Variability in reliability coefficients and the standard error of measurement from school district to district. *Applied Measurement in Education*, *12*, 367-381.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York, NY: McGraw-Hill.

Fine, E. M., & Delis, D. C. (2011). Delis-Kaplan Executive Functioning System. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 796-801). New York, NY: Springer.

Floyd, R. G., McCormack, A. C., Ingram, E. L., Davis, A. E., Bergeron, R., & Hamilton, G. (2006). Relations between the Woodcock-Johnson III Clinical Clusters and measures of executive functions from the Delis-Kaplan Executive Function System. *Journal of Psychoeducational Assessment*, *24*, 303-317.

Folstein, M. F., Folstein, S. E., & McHugh. P. R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinicians. *Journal of Psychiatry Research*, *12*, 189-198.

Ghawami, H., Sadeghi, S., Raghibi, M., & Rahimi-Movaghar, V. (2016). *Executive functioning of complicated-mild to moderate traumatic brain injury patients with frontal contusions*. *Applied Neuropsychology: Adult*. doi:10.1080/23279095.2016.1157078

Gillespie, D. C., Evans, R. I., Gardener, E. A., & Bowen, A. (2002). Performance of older adults on tests of cognitive estimation. *Journal of Clinical and Experimental Neuropsychology*, *24*, 286-293.

Gouveia, P. A. R., Brucki, S. M. D., Malheiros, S. M. F., & Bueno, O. F. A. (2007). Disorders in planning and strategy application in frontal lobe lesion patients. *Brain and Cognition*, *63*, 240-246.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Test Technology, 1,* 1-12.

Hancock, M., Tapscott, J. L., & Hoaken, P. N. S. (2010). Role of executive dysfunction in predicting frequency and severity of violence. *Aggressive Behavior*, *36*, 338-349.

Homack, S., Lee, D., & Riccio, C. A. (2005). Test review: Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*, *27*, 599-609.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health*, *7*, 131-142.

Jefferson, A. L., Paul, R. H., Ozonoff, A., & Cohen, R. A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living (IADLs). *Archives of Clinical Neuropsychology*, *21*, 311-320.

Jones-Chesters, M. (2008). *D-KEFS validity: An update of the research*. Oxford: Pearson Assessment. Retrieved from: http://roar.uel.ac.uk/1512/

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll, & B. K. Bryant (Eds.), *Clinical neuropsychology & brain function: Research, measurement & practice* (pp 127-167). Washington, DC: American Psychological Association.

Keifer, E., & Tranel, D. (2013). A neuropsychological investigation of the Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology*, *35*, 1048-1059.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press.

Lovstad, M., Fundetud, I., Endestad, T., Due-Tonnessen, P., Meling, T. R., Lindgren, M., & Solbakk, A. K. (2012). Executive functions after orbital or lateral prefrontal lesions: Neuropsychological profiles and self-reported executive functions in everyday living. *Brain Injury*, *26*, 1586-1598.

Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, *50*, 451-468.

Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is use of the wooden and the computerized versions of the Tower of Hanoi puzzle equivalent? *Applied Neuropsychology*, *9*, 117-120.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties*. New York, NY: Routledge.

Mitchell, M., & Miller, L. S. (2008). Prediction of functional status in older adults: The ecological validity of four Delis-Kaplan Executive Function System tests. *Journal of Clinical & Experimental Neuropsychology*, *30*, 683-690.

Mitrushina, M., Boone, K., Razani, J., & D'Elia, L. (2005). *Handbook of Normative Data for Neuropsychological Assessment* (2nd ed.). New York, NY: Oxford University Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice*. Mahwah, NJ: Erlbaum.

Norris, G. & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation*, *10*, 33-45.

Ohlman, H. (1959). Subject-Word letter frequencies with applications to superimposed coding. In National Academy of Sciences*, Proceedings of the International Conference on Scientific Information* (pp. 903-916). Retrieved from: http://www.nap.edu/catalog/10866.html

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *The Clinical Neuropsychologist*, *27*, 1328-1361.

Phillips, L. H. (1997). Do "frontal tests" measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 191-213). Hove: Psychology Press.

Poortinga, Y. H. (1989). Equivalence of cross cultural data: An overview of basic issues. *International Journal of Psychology*, *24*, 737-756.

Puente, A. E., & Perez-Garcia, M. (2000). Psychological assessment of ethnic minorities. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (3rd ed.) (pp. 527-552). New York, NY: Pergamon.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1-38). Hove: Psychology Press.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*, 566-594.

Schmidt, M. (2003). Hit or miss? Insight into executive functions: Delis-Kaplan Executive Functions System [Book review]. *Journal of the International Neuropsychological Society*, *9*, 962-964.

Seyedian, M., Falah, M., Noroozyan, M., Nejat, S., Delavar, A., & Ghasemzadeh, H. (2007). Tahie va taiine etebere noskheye farsiye azmoone kutahe vazeiate zehni [Development and validation of Persian version of Mini-Mental State Examination]. *Journal of Medical Council of Islamic Republic of Iran*, *25*, 408-414.

Shunk, A. W., Davis, A. S., & Dean, R. S. (2006). Test review: Delis-Kaplan Executive Function System. *Applied Neuropsychology*, *13*, 275-279.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York, NY: Oxford University Press.

Swanson, J. (2005). The Delis-Kaplan Executive Function System: A review. *Canadian Journal of School Psychology*, *20*, 117-128.

Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.

Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*, 119-135.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). Behavioural Assessment of the Dysexecutive Syndrome (BADS). London: Thames Valley Test Company.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a Dysexecutive Syndrome. *Neuropsychological Rehabilitation, 8*, 213-228.

World Health Organization. (2009). *Global status report on road safety: Time for action*. Geneva: Author.

World Health Organization. (2013). *Global status report on road safety: Supporting a decade of action*. Geneva: Author.

Yochim, B. P., Baldo, J. V., Kane, K. D., & Delis, D. C. (2009). D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: The importance of error monitoring. *Journal of Clinical and Experimental Neuropsychology*, *31*, 658-663.

RECEIVED: October 26, 2015
ACCEPTED: June 18, 2016