ORIGINAL ARTICLE
https://doi.org/10.4306/pi.2017.14.2.126

Print ISSN 1738-3684 / On-line ISSN 1976-3026
OPEN ACCESS

Effects of Education on Differential Item Functioning on the 15-Item Modified Korean Version of the Boston Naming Test

Byung-Soo Kim[1], Dong-Woo Lee[2], Jae-Nam Bae[3], Ji-Hyun Kim[3], Shinkyum Kim[4], Ki Woong Kim[5,6], Jee-Eun Park[7], Maeng Je Cho[6,7], and Sung Man Chang[1] ✉

[1]Department of Psychiatry, Kyungpook National University School of Medicine, Daegu, Republic of Korea
[2]Department of Psychiatry, Inje University College of Medicine, Seoul, Republic of Korea
[3]Department of Psychiatry, Inha University College of Medicine, Incheon, Republic of Korea
[4]Department of Psychiatry, Yangsan Mental Hospital, Yangsan, Republic of Korea
[5]Department of Neuropsychiatry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[6]Department of Psychiatry and Behavioral Science, Seoul National University College of Medicine, Seoul, Republic of Korea
[7]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea

Objective Education is expected to have an effect on differential item functioning (DIF) on the 15-item Modified Boston Naming Test in the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (BNT-KC). However, no study has examined DIF in the BNT-KC.

Methods We used the item response theory to investigate the impact of education on the DIF in the BNT-KC among elderly individuals with or without dementia (n=720). A two-parameter item response model was used to determine the difficulty and discrimination parameters of each item. The Benjamini-Hochberg procedure was used to address the risk of Type I errors on multiple testing.

Results Four items, "mermaid," "acorn," "compass," and "pomegranate" continued to demonstrate DIF after controlling for multiple comparisons. Those with low education levels were more likely to error on "mermaid" and "compass," while those with high education levels were more likely to error on "acorn" and "pomegranate." "Hand" and "red pepper" were too easily identified to be used for detecting dementia patients. "Monk's hat" and "pomegranate" were less discriminating than other items, limiting their usefulness in clinical setting.

Conclusion These findings may provide useful information for the development of a revised version of the BNT-KC to help clinicians make diagnostic decisions more accurately.
Psychiatry Investig 2017;14(2):126-135

Key Words Language tests, Dementia, Education, Diagnostic errors.

INTRODUCTION

The Boston Naming Test (BNT), introduced in 1983 by Kaplan, Goolglass, and Weintraub, may be the most widely used naming test worldwide.[1] Impairment in language function is one of the core symptoms of dementia,[2] so the BNT has been used in clinical situations to support the diagnosis of dementia. Moreover, BNT performance is useful for predicting the subsequent development of AD in preclinical cases or prognosis in patients with early AD.[3,4]

Meanwhile, the validity of tools used to assess cognitive function is usually influenced by a subject's educational attainment.[5] Specifically, those with low education levels are often embarrassed and sometimes refuse to complete pencil-and-paper-based assessments. For those who are not accustomed to reading and writing, confrontational naming tests like the BNT may be more comfortable to perform because literacy is not required. From this point of view, the BNT is an especially useful tool for evaluating cognitive function in elderly Korean individuals since many have been deprived of the chance to learn to write during the Japanese colonization period and Koran war.

A 60-item Korean version of the BNT (K-BNT) was standardized by Kim and Na.[6] The K-BNT uses a similar administration method to that of the original BNT, but most of the items were changed due to the linguistic and cultural differences between Korean and English speakers.[7] When admin-

Received: December 25, 2015 **Revised:** February 24, 2016
Accepted: March 31, 2016 **Available online:** October 13, 2016

✉ **Correspondence:** Sung Man Chang, MD, PhD
Department of Psychiatry, Kyungpook National University School of Medicine, 130 Dongdeok-ro, Jung-gu, Daegu 41944, Republic of Korea
Tel: +82-53-420-5753, **Fax:** +82-53-426-5361, **E-mail:** psyjang@hanmail.net

istered to patients with dementia, the length of the 60-item version often detracts from the patient's concentration and can be a barrier to completion. To address this limitation, several abbreviated forms have been proposed.[8,9] The 15-item Modified Boston Naming Test (BNT-KC) in the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K) Neuropsychological Assessment Battery (CERAD-K-N) is one of them.[10] Fifteen items were chosen from the K-BNT. Three groups of five items each were identified as having high, medium, and low frequencies of occurrence. Phonemic and semantic dissimilarities in arranging the items were also considered.[10] BNT-KC is the most widely used short version of the K-BNT because it is included in the CERAD-K-N, the most extensively used neuropsychological test battery in Korea.

Because cognitive test results are influenced by demographic factors such as age, sex, and educational level, normative information on eight tests in the CERAD-K-N were provided to help interpret the results. However, normative data does not give us enough information about the differences between the items. If the conditional probability of a correct response differs between two ability-matched groups, the item demonstrates differential item functioning (DIF).[11] An item with DIF measures differently for one subgroup than another, resulting in less subgroup assessment validity. Thus, understanding the DIF that leads to intergroup score discrepancies may have important clinical implications that normative data cannot provide.

To date there have been several reports about DIF in the Mini-Mental State Examination (MMSE),[12-14] but there has been little research about DIF in the BNT except for race/ethnicity attributes.[11] Considering that educational attainment in Korean elderly individuals varies widely from illiteracy to high education levels, DIF in BNT according to educational level is strongly expected. For example, some items of the BNT-KC such as "mermaid" and "dinosaur" are not usually spoken in ordinary life and are encountered only in literature or movies. Thus, individuals with low education levels, especially those who are illiterate, may have more difficulty responding correctly to these items despite not having naming ability impairments.

This study aimed to investigate the impact of education on DIF in BNT-KC in a clinical sample. To address this issue, the current study used the item response theory (IRT), which was used previously to study DIF in various psychiatric assessment tools.[15-17]

METHODS

Sample

The study sample included total 773 elderly individuals, 386 from the National Dementia Epidemiological Study in Korea (NDESK) and 387 from the memory clinic of Kyungpook National University Medical Center (KNUMC). The study protocol for the NDESK was previously described in detail.[18] In brief, the NDESK study investigated the prevalence and associated risk factors of dementia from a nationwide sample of Korean elders. A multi-stage cluster sampling design was adopted. Fifteen catchment areas were selected across six major administrative districts, including metropolitan Seoul and five provinces. Approximately 500 participants (65 years or older) were allotted to each catchment area. A total of 8,199 elderly individuals were invited to participate in the Phase I screening assessment using the Korean version of the MMSE (MMSE-KC) by door-to-door home visits; among them, 6,141 subjects responded (response rate=74.9%). Of those subjects, 2,336 were invited to participate in the Phase II diagnostic assessment for dementia and 1,673 subjects responded (response rate=71.6%). The CERAD-K-N was used to assess cognitive impairment.[10] Among the 15 catchment areas, five catchment areas (Yeoncheon-gun, Incheon, Daegu, Gwangju, and Jungnang-gu in Seoul) were selected to investigate the prevalence of sleep problems and depression with executive dysfunction (DED). A total of 3,074 participants were selected from the five catchment areas, of which 2,007 subjects completed Phase I interviews (response rate=65.3%). Of those subjects, 806 were invited to participate in the Phase II diagnostic assessment for dementia, of which 557 subjects responded (response rate=69.1%). Data from 386 participants from five research centers who completed the BNT-KC were used in the current study. The study sample from the memory clinic in KNUMC included 387 elderly individuals aged ≥60 years who were evaluated for the presence of dementia between April 2013 and February 2015. The Institutional Review Board of KNUMC (Daegu, Korea) approved this study. From both samples, the individuals who were unable to complete the BNT-KC due to a visual disturbance, hearing difficulty, poor general condition, or other psychiatric condition were excluded. Among the 773 elderly who completed the BNT-KC, the participants who scored <10 on the MMSE-KC (n=53; 22 from NDESK and 31 from KNUMC) were also excluded because severe cognitive impairments can prevent proper test performance. Ultimately, the data of 720 individuals were analyzed (364 from NDESK and 356 from KNUMC).

Assessments

The CERAD-K-N was administered to every participant by

neuropsychologists or trained research nurses.[18] The CERAD-K-N consists of nine neuropsychological tests as follows: Verbal Fluency Test, 15-item Modified Boston Naming Test, MMSE-KC, Word List Memory Test, Constructional Praxis Test, Word List Recall Test, Word List Recognition Test, Constructional Recall Test, and Trail Making Test.[10] All instruments were previously validated in the Korean population.[19] The BNT-KC was administered to participants without cueing. All 15 line drawings were presented and the total number of spontaneous correct responses was calculated. In both NDESK and KNUMC samples, the diagnosis of dementia was defined according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition diagnostic criteria after a thorough clinical interview by clinicians, laboratory tests including complete blood cell count, chemistry profile, serological test for syphilis, echocardiography, chest X-ray, and brain computed tomography or magnetic resonance imaging.[2]

Statistical analysis

IRT model

To evaluate DIF in BNT-KC between the low and high education groups, a two-parameter item response model was used. The basic goal of IRT modeling is to determine item response functioning (IRF) for each item. An IRF is a mathematical function that describes the link between where an individual falls on the continuum of a given construct such as naming ability and the probability that he or she will give a particular response to a scale item designed to measure that construct, and it can be used to evaluate item quality.[20] Difficulty is the ability value that is associated with a 50% probability of responding correctly on an individual item.[21] *Discrimination* is an index of how well an item can differentiate between patients of varying severity levels.[21] Item difficulty and discrimination are easily visualized using an item characteristic curve (ICC) for each item (Figure 1). Item difficulty is the location along the latent trait axis (x-axis) where the IRF curve changes direction (its inflection point; β in Figure 1).[20] Item discrimination is the steepness of the IRF at the curve's inflection point (i.e., at the item's difficulty level; α in Figure 1).[20] A two-parameter IRT model estimates a difficulty parameter and a discrimination parameter of the item.

Unidimensionality assumption

The important assumption of IRT is unidimensionality, which means that a single latent trait is sufficient to account for the observed pattern of item responses. The BNT-KC theoretically assumes that responses to each item are linked to a single construct of naming ability. Statistically, we tested the unidimensionality of the BNT-KC with confirmatory factor analysis using AMOS 21 (2012; IBM, Armonk, NY, USA). Model fit was evaluated with the comparative fit index (CFI), Tucker-Lewis Index (TLI), and the root-mean-square error of approximation (RMSEA).

Differential item functioning

We used IRTLRDIF version 2.0b (D. Thissen, IRTLRDIF v. 2.0b: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning, 2001, unpublished manuscript) to carry out the DIF analyses.[22] IRTLRDIF uses the likelihood-ratio test statistic to provide a significance test for the null hypothesis that the parameters of an item's response function do not differ between groups, and the significance of that test statistic is the detection of DIF.[22]

In the current study, the null hypothesis was that the item

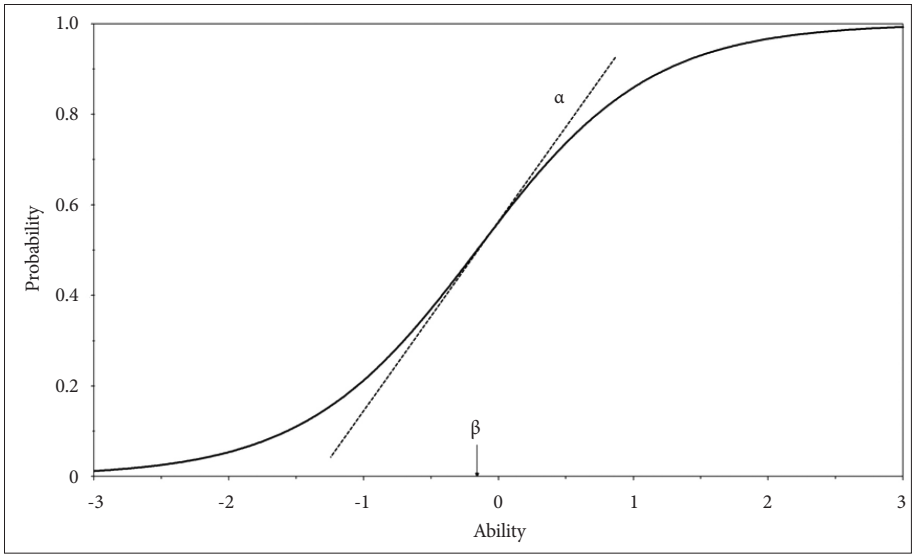

Figure 1. Item characteristic curve. α: discrimination parameter, β: difficulty parameter.

parameters do not differ between the low and high education groups. Analyses proceeded by initially constraining both difficulty and discrimination estimates were equal for the two groups across all 15 items. This step produces $ll_{AllEqual}$, the log likelihood for all item parameters constrained equal (no DIF for any item). For each of the 15 items, a model was then fit that constrains all of the remaining items' difficulty and discrimination estimates to be equal but allows the estimates for one item (Item I) to differ between the two groups ($ll_{ItemInotEqual}$). The value of G^2 (d.f.)=-2 ($ll_{AllEqual}$-$ll_{ItemInotEqual}$) provides an omnibus test (df=2) of whether there is DIF for the difficulty and/or discrimination estimate for this item. The difference between the log-likelihood statistics of the two models is distributed as a χ^2 statistic (G^2) with degrees of freedom equal to the difference in parameter estimates between the two models.[22] A significant G^2 statistic at the $p<0.05$ level indicates that at least one of the parameters differs between the groups and is assumed to demonstrate DIF. Given that we conducted DIF analyses across multiple items, it is important to account for the risk of Type I errors. We employed the Benjamini-Hochberg procedure because it has been shown to demonstrate greater power than Bonferroni correction.[23,24]

RESULTS

Sample characteristics

Characteristics of the final sample are summarized in Table 1. Participants were grouped into high (≥7 years) and low (≤6 years) education groups. The proportion of women in the low education group was higher than that in the high education group (71.2% vs. 34.8%, χ^2=88.294, p<0.001). The low education group was significantly older (74.7±6.6 vs. 72.5±6.1, t=4.306, df=718, p<0.001) and had lower MMSE-KC scores (19.7±4.8 vs. 23.8±4.4, t=-11.480, df=487.327, p<0.001). The proportion of individuals living in rural areas (Yeoncheon-gun) was higher in the low education group compared to the high

Table 1. Characteristics of the study participants (N=720)

Participant characteristic	All participants (N=720)	Low education group (N=490)	High education group (N=230)	p value*
	N (%) or mean±SD	N (%) or mean±SD	N (%) or mean±SD	
Age at baseline interview (years)				0.004
60–64	31 (4.3)	18 (3.7)	13 (5.7)	
65–69	162 (22.5)	97 (19.8)	65 (28.3)	
70–74	205 (28.5)	134 (27.3)	71 (30.9)	
75–79	178 (24.7)	127 (25.9)	51 (22.2)	
80–84	91 (12.6)	69 (14.1)	22 (9.6)	
≥85	53 (7.4)	45 (9.2)	8 (3.5)	
Sex				<0.001
Female	428 (59.4)	141 (28.8)	151 (65.7)	
Male	292 (40.6)	349 (71.2)	79 (34.3)	
Diagnosis				0.274
Normal	452 (62.8)	301 (61.4)	151 (65.7)	
Dementia	268 (37.2)	189 (38.6)	79 (34.3)	
Sampling site				0.450
NDESK (community)	356 (49.4)	243 (49.6)	121 (52.6)	
KNUMC (hospital)	364 (50.6)	247 (50.4)	109 (47.4)	
Residence				0.006
Urban	570 (79.2)	374 (76.3)	196 (85.2)	
Rural	150 (20.8)	116 (23.7)	34 (14.8)	
Years of education	5.6±4.6	3.0±2.7	11.0±2.8	<0.001
Age at baseline interview (years)	74.0±6.6	74.7±6.6	72.5±6.1	<0.001
MMSE-KC scores	21.0±5.1	19.7±4.8	23.8±4.4	<0.001

*chi-square tests were used for categorical variables and t-tests were used for continuous variables. NDESK: National Dementia Epidemiological Study in Korea, KNUMC: Kyungpook National University Medical Center, SD: standard deviation, MMSE-KC: Korean version of the Mini-Mental State Examination

education group (23.7% vs. 14.8%, χ^2=7.502, p=0.006). However, there was no significant difference in the proportion of individuals who were diagnosed with dementia between the low and high education groups (38.6% vs. 34.3%, χ^2=1.195, p=0.274).

Unidimensionality assumption

Results from a confirmatory factor analyses showed χ^2=185.972, CFI=0.912, TLI=0.897, and RMSEA=0.047. These findings indicated a reasonable data fit. Results from a confirmatory factor analyses showed χ^2=107.701, CFI=0.936, TLI=0.924, and RMSEA=0.042. These findings indicated a reasonable data fit. We determined that these fit statistics were sufficient to proceed to fitting IRT models.

Differential item functioning

Table 2 shows the frequency of incorrect answers to each item of the BNT-KC in the low and high education groups as well as corresponding difficulty and discrimination parameter estimates. The ICC for each item is demonstrated in Figure 2. For every item, the frequency of incorrect answers was higher in the low education group than in the high education group, which was reflected in the difference in mean BNT-KC scores (7.5±3.1 vs. 10.0±2.9, t=-10.766, df=718, p<0.001). Seven of 15 items initially demonstrated DIF: "balloon," "bat," "mermaid," "acorn," "compass," "pomegranate," and "monk's hat." After controlling for multiple comparisons, four items, namely "mermaid," "acorn," "compass," and "pomegranate," continued to demonstrate DIF, while the other items did not. Among the four items demonstrating DIF, "acorn" and "pomegranate" showed higher β parameters in the low education group, whereas "mermaid" and "compass" showed higher β parameters in the high education group.

The order of the four most difficult items was the same in both groups: "monk's hat" (β=1.54 for the low education group, β=1.65 for the high education group), "wind bell" (β=1.10 for the low education group, β=0.84 for the high education group), "dinosaur" (β=0.54 for the low education group, β=0.66 for the high education group), and "funnel" (β=0.14 for the low education group, β=0.21 for the high education group). "Pomegranate" is usually considered a difficult item because it is a low frequency word in the BNT-KC. However, its difficulty parameter (β=-0.77) was similar to those of middle frequency words (β=-0.87−-0.64) in the high education group. Moreover, it was even lower (β=-1.42) than those of the middle frequency words (β=-1.18−-0.23) in the low education group.

The order of the five easiest items was the same in both education groups: "hand" (β=-4.45 for the low education group, β=-151.95 for the high education group), "red pepper" (β=-3.96 for the low education group, β=-9.16 for the high education group), "baduk (go)" (β=-2.35 for the low education group, β=-2.54 for the high education group), "cobweb" (β=-1.96 for the low education group, β=-2.15 for the high education group), and "balloon" (β=-1.80 for the low education

Table 2. Item parameters for study participants (N=720)

Item frequency	Item no.	Name	Low education group			High education group			G^2 (df=2)
			Incorrect answer (%)	α	β	Incorrect answer (%)	α	β	
Low	1	Hand	0.8	1.86	-4.45	0.0	0.29	-151.95	0.6
	2	Red pepper	1.8	1.85	-3.96	0.4	0.61	-9.16	1.2
	3	Baduk (go)	14.5	2.18	-2.35	2.2	2.54	-2.54	2.0
	4	Cobweb	24.1	1.82	-1.96	7.8	1.54	-2.15	0.4
	5	Balloon	30.1	1.49	-1.80	17.0	2.27	-1.16	9.2*†
Middle	6	Bat	45.5	2.34	-1.15	24.3	2.06	-0.87	7.2*†
	7	Traffic light	60.1	1.75	-0.65	30.9	1.49	-0.74	0.8
	8	Mermaid	68.2	2.54	-0.42	27.0	2.30	-0.79	10.7**
	9	Acorn	45.3	1.66	-1.18	31.3	1.73	-0.64	13.7**
	10	Compass	72.7	2.19	-0.23	31.3	1.57	-0.74	17.6***
High	11	Dinosaur	88.0	1.92	0.54	66.5	1.49	0.66	1.1
	12	Pomegranate	42.4	0.87	-1.42	38.3	0.66	-0.77	10.6**
	13	Funnel	75.9	1.27	0.14	55.7	1.87	0.21	4.2
	14	Wind bell	91.8	1.52	1.10	72.6	1.74	0.84	1.0
	15	Monk's hat	85.9	0.78	1.54	82.6	1.19	1.65	8.1*†

*p<0.05, **p<0.01, ***p<0.001, †differential item functioning test not statistically significant after Benjamini-Hochberg procedure, G^2=chi-square with two degrees of freedom (df)

group, β=-1.16 for the high education group), all of which were high frequency words.

In the low education group, the least discriminating item was "monk's hat" (α=0.78), which represented the most difficult item. The second-least discriminating item was "pomegranate" (α=0.87) in this group. In the high education group, the least discriminating item was "hand" (α=0.29) and the second-least discriminating item was "red pepper" (α=0.61), which represented the easiest items, and the error rate was <2% in both groups. Interestingly, the third-least discriminating item in the high education group was "pomegranate" (α=0.66), which was the second least discriminating item in the low education group. The most highly discriminating item in the low education group was "mermaid" (α=2.54), while that in the high education group was "baduk (go)" (α=2.54).

DISCUSSION

Here we explored the influence of education on DIF in BNT-KC using IRT. To the best of our knowledge, this was the first study to investigate DIF of BNT by educational attainment. To date, few studies have examined DIF in the BNT except with respect to race/ethnicity attributes.[11] In contrast, several studies have reported the influence of education on DIF in the MMSE.[12-14,25] These studies have reported that naming items (e.g., pencil, watch, etc.) are biased by level of education;[14,25] however, this finding is not consistent across studies.[13]

In the current study, we found DIF in several items of the BNT-KC that were attributable to educational attainment. Those with low education levels were more likely to respond incorrectly to "mermaid" and "compass" despite having comparable naming ability to that of individuals with high education levels, whereas those with high education levels experienced greater difficulty correctly responding to "acorn" and "pomegranate" compared to those with low education levels.

Both "mermaid" and "compass" are items that are uncommon in everyday life. "Mermaid" is a character in fairytales of western culture of which most Korean elders with low education levels are ignorant. Wrong answers often include "half girl, half fish" or "fish woman," and even after they were explained the answer (mermaid), they did not understand it. Considering that aged people in Korea had a lower chance of being familiar with western culture, the cultural factors should be considered in the process of developing a revised naming test to aid the diagnosis of dementia.

"Compass" is not a familiar item in everyday life as well. Most people have a basic understanding of a compass—its magnetic characteristic and the usage for orientation—from their time in school. Thus, the higher difficulty experienced in identifying this item in the low educated group is not surprising. Another possibility is that this result may come from the confounding effect of sex on education since, in the current sample, the high education group showed a higher proportion of male sex than the low education group. Korea has a conscription system for men, who are taught how to read a map using a compass during military life. In the analysis of DIF using IRTLRDIF, we could not control the confounding effect of other variables on the results. To eliminate the confounding effect of sex, we had to match the sample according to sex, but we did not consider it in the study design because the smaller sample size may decrease the detectable effect size at a given power.[26] To address this limitation, a larger study based on a sex-matched sample may be required.

Two items, "acorn" and "pomegranate," demonstrated lower naming difficulty in the low education group than in the high education group. Decades ago, individuals raised in rural areas usually stopped going to school after elementary school to help with the farm work. Thus, those individuals usually had fewer years of education than those raised in urban areas. However, both "acorn" and "pomegranate" are fruits of plants that can be more commonly seen in rural than in urban areas. This may have contributed to the lower naming difficulty of these items in the low education group. The finding in the current study that the proportion of individuals living in rural areas was higher in the low education group compared to the high education group may support this interpretation. However, given that individuals may learn the words "acorn" and "pomegranate" during childhood or adolescence, data about childhood residence may be more critical than data about current residence. Unfortunately, we did not collect data on childhood residence in this study.

It is important to note that this result does not mean that the proportion of individuals who responded correctly to "acorn" and "pomegranate" was higher in the low education group than in the high education group. Rather, it means that if the individuals have the same naming ability (i.e., total score on the BNT-KC), the probability of responding correctly to "acorn" and "pomegranate" is higher in individuals with low education levels than in those with high education levels. Although this simplified explanation assumes that DIF types are uniform,[11] all of the items in the current study demonstrated uniform DIF (Figure 2).

It is usually expected that the less frequent the word is encountered, the more difficult it is for one to name it. However, in case of "pomegranate," even though it is a low frequency word, its naming difficulty was similar to those of middle frequency words. This demonstrates that it is not always true that the frequency of exposure to a word is inversely correlated with difficulty naming it.

"Hand" and "red pepper" were too easy for most subjects, so

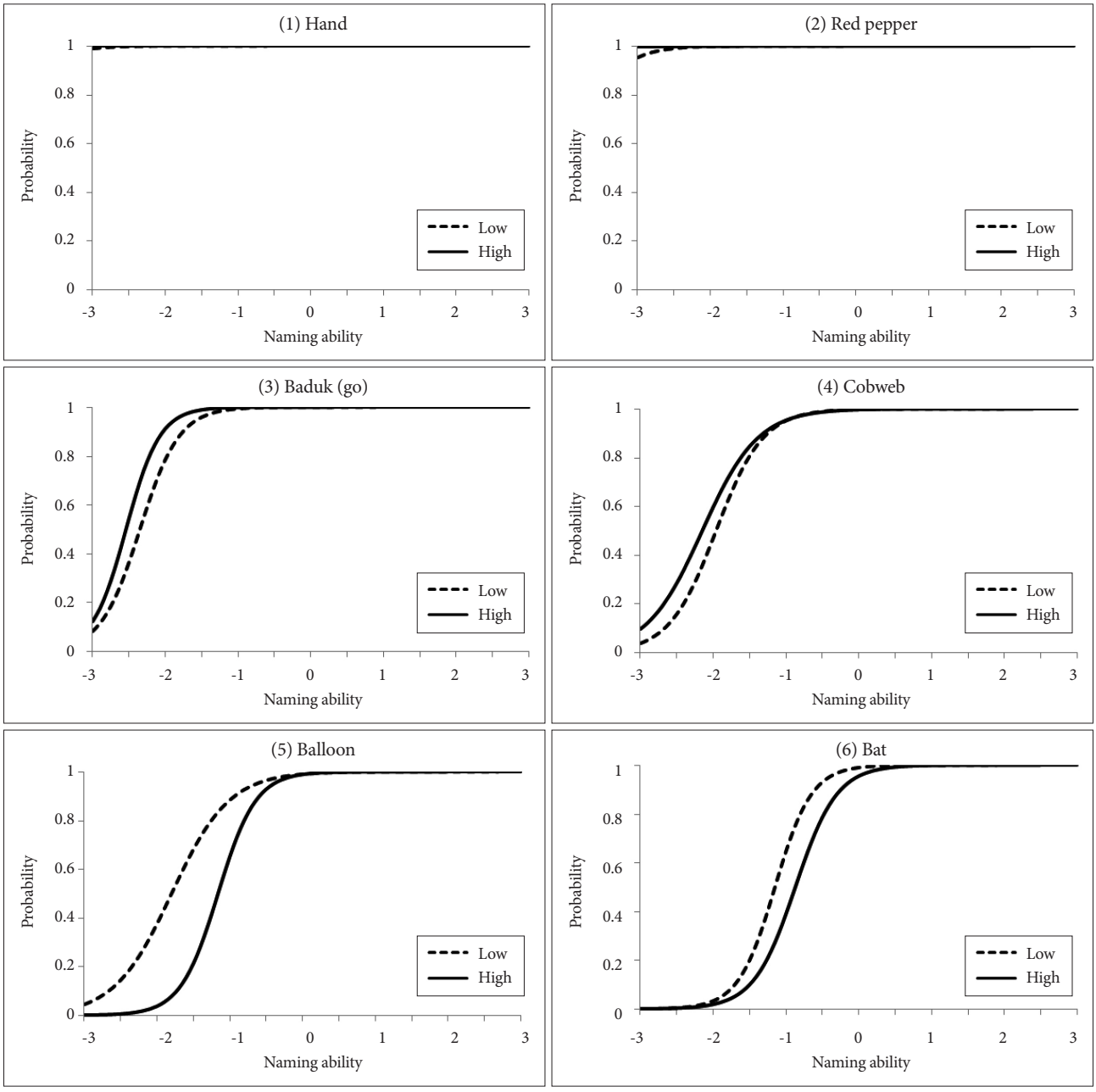

Figure 2. Item characteristic curves for the items of the 15-item Modified Boston Naming Test in the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet Neuropsychological Assessment Battery (BNT-KC). Four items, namely "(8) mermaid", "(9) acorn", "(10) compass", and "(12) pomegranate", demonstrated differential item functioning between the low and high education groups. "(1) Hand" and "(2) red pepper" were too easy for most subjects. "(12) Pomegranate" and "(15) monk's hat" seemed less useful for the diagnosis of dementia due to their low discriminating feature.
(Continued to the next page)

these items appeared less useful for the early diagnosis of dementia in the clinical setting. Besides, "pomegranate" and "monk's hat" seemed less useful for the diagnosis of dementia due to their low discriminating feature. In other words, many individuals with a normal naming ability failed to correctly name these items, while persons with impaired naming ability frequently named them correctly. The low discriminating feature of these items may in part be attributable to the vagueness of the figure in the BNT-KC. Persons with normal cognitive function often mistake the pomegranate for a flower and the monk's hat for a towel. Further study of the visual perceptual errors of these items is needed.

We must acknowledge several study limitations. First, as mentioned earlier, we did not control for the possible confounding effects of other variables such as age and sex. We did not observe an interaction between education and potential confounding variables such as age, sex, and residence. Nevertheless, differences in sociodemographic variables be-

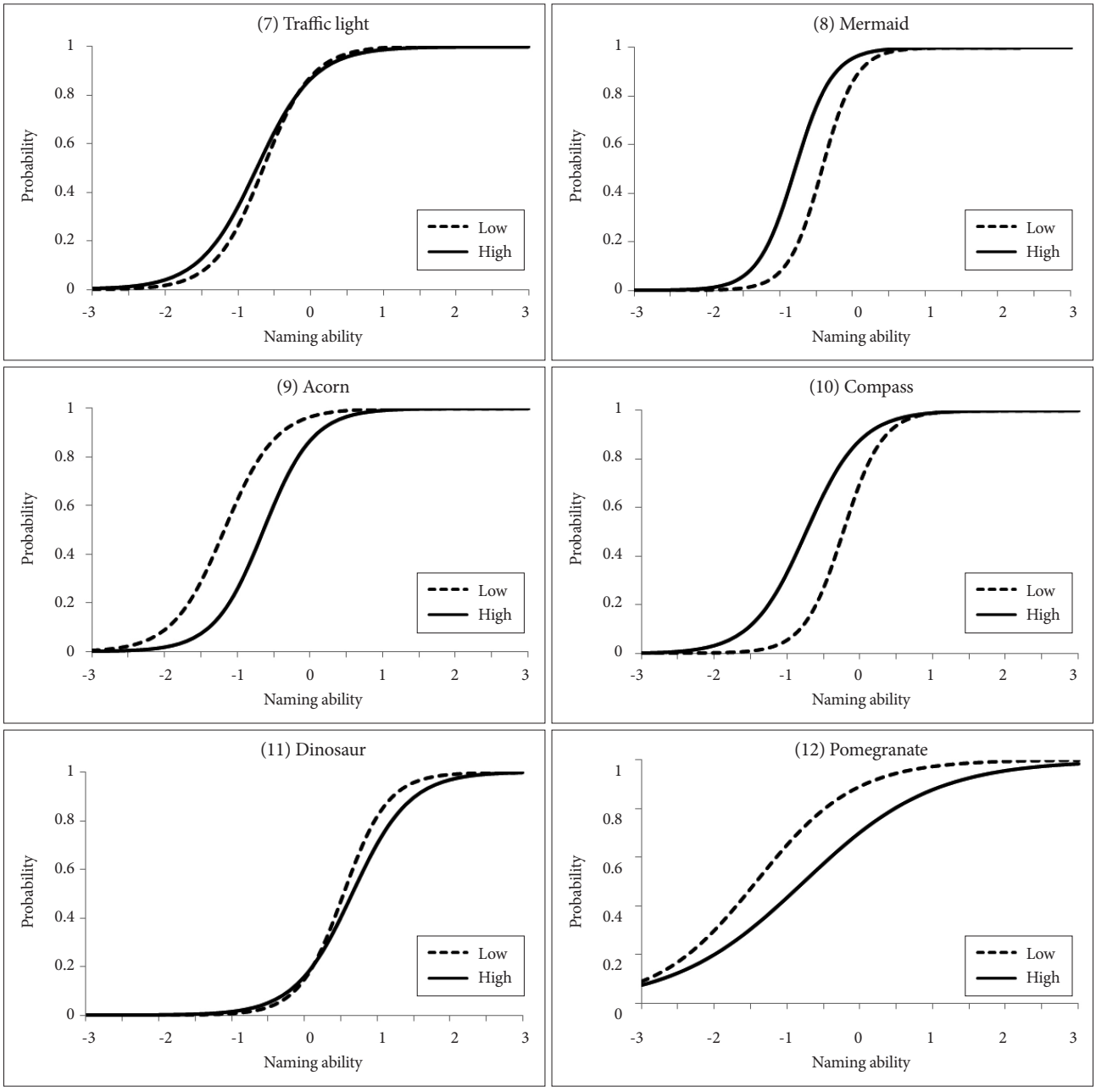

Figure 2. (Continued form the previous page) Item characteristic curves for the items of the 15-item Modified Boston Naming Test in the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet Neuropsychological Assessment Battery (BNT-KC). Four items, namely "(8) mermaid", "(9) acorn", "(10) compass", and "(12) pomegranate", demonstrated differential item functioning between the low and high education groups. "(1) Hand" and "(2) red pepper" were too easy for most subjects. "(12) Pomegranate" and "(15) monk's hat" seemed less useful for the diagnosis of dementia due to their low discriminating feature. (Continued to the next page).

tween low and high education groups may have influenced the findings of this study. Several studies have reported an association between increasing age and lower BNT scores.[19,27,28] In the current sample, the low education group was approximately 2 years older than the high education group. However, this difference was not large enough to infer any clinical significance, given that normative values for the BNT are usually presented in at least 5-year intervals. The influence of sex on language remains controversial. Although sex did not affect scores on the original BNT or 60-item K-BNT,[7,27] sex had a significant effect on the total BNT-KC score, necessitating the development of normative values for age, sex, and education level.[19] Several investigators have reported that AD has a greater negative impact on language function in female patients,[29,30] although this finding is not consistent across studies.[31] To the best of our knowledge, no previous reports have found an influence of sex on DIF in the BNT. However, one study found an influence of sex on DIF in the Picture Nam-

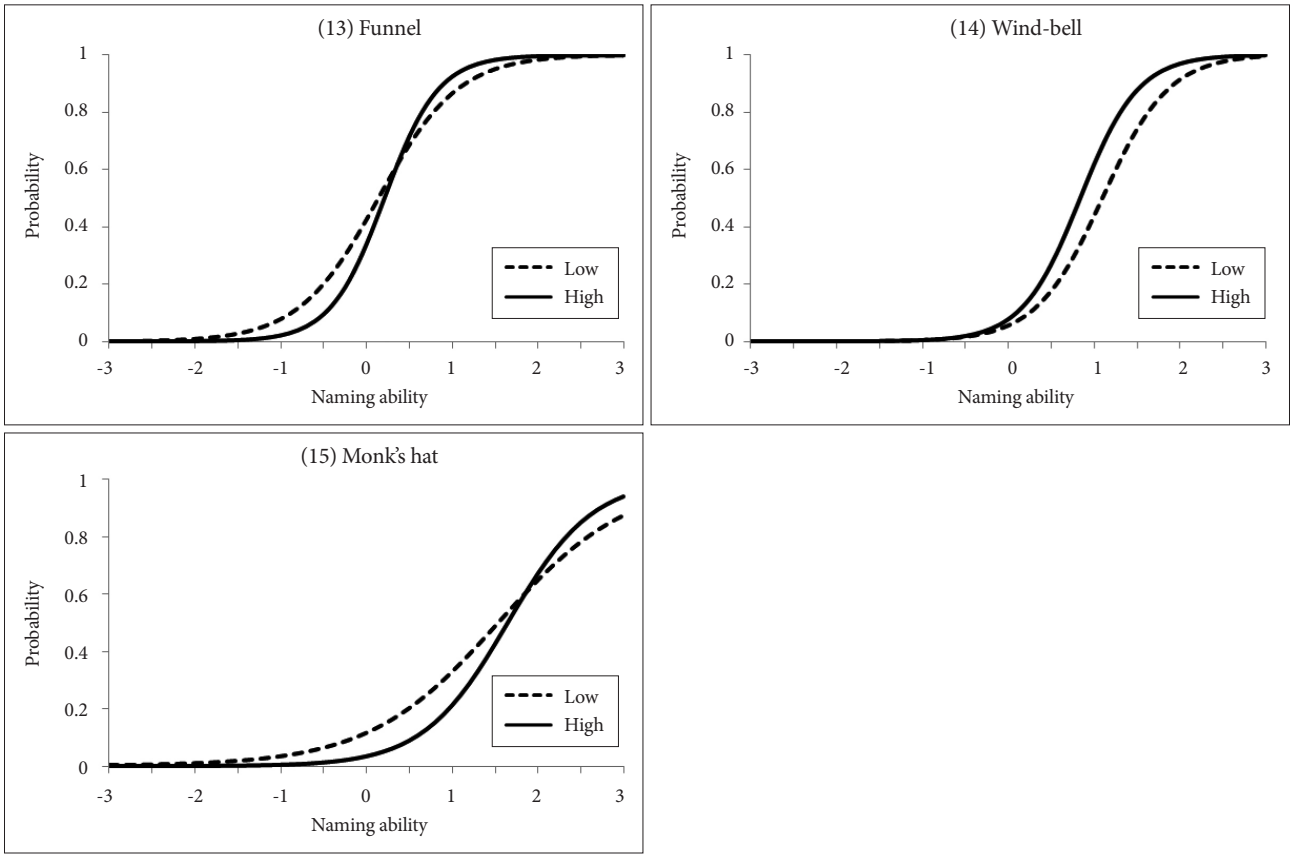

Figure 2. (Continued form the previous page) Item characteristic curves for the items of the 15-item Modified Boston Naming Test in the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet Neuropsychological Assessment Battery (BNT-KC). Four items, namely "(8) mermaid", "(9) acorn", "(10) compass", and "(12) pomegranate", demonstrated differential item functioning between the low and high education groups. "(1) Hand" and "(2) red pepper" were too easy for most subjects. "(12) Pomegranate" and "(15) monk's hat" seemed less useful for the diagnosis of dementia due to their low discriminating feature.

ing in Repeatable Battery Assessment of Neuropsychological Status (RBANS),[32] which suggests that sex is a potential confounding factor in the current study. The only method we know to control for confounding in DIF analyses based on IRT is matching; however, we were unable to match samples in the current study due to our limited sample size. To adjust for multiple variables without matching, other statistical methods such as the multiple indicators-multiple causes (MIMIC) model or logistic regression model can be used.[13,33] However, we employed the IRT model to explore item characteristics (e.g., difficulty, discrimination, etc.), which cannot be measured by other approaches. In the future, a sufficiently large sample matched for age and sex may be warranted to examine the effect of education on DIF. Second, patients with dementia were considered a homogeneous group in the current study. However, there have been reports that there are differences in naming difficulty patterns among dementia types (i.e., AD vs. vascular dementia; AD vs. frontotemporal dementia).[34,35] Although the proportion of patients with dementia did not differ between the low and high education groups, the different dementia types could be influential. Third, the results of this study cannot be generalized to other countries using different language versions of the BNT since the items in the BNT-KC were altered from the original BNT to compensate for the linguistic and cultural differences between Korean and English speakers.[7] Nevertheless, the results of this study suggest that DIF attributable to education could exist in other language versions of the BNT.

In summary, the current study revealed that elderly individuals in Korea with low education levels are more likely to incorrectly name "mermaid" and "compass," while those with high education levels are more likely to incorrectly name "acorn" and "pomegranate." "Hand" and "red pepper" are too easily named for use in the early detection of dementia, while "monk's hat" and "pomegranate" are less discriminating than other items, which limits their usefulness in the clinical setting. These findings may provide useful information for the development of a revised version of the BNT-KC to help clinicians make diagnostic decisions more accurately.

REFERENCES

1. Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. Phila-

delphia: Lea & Febiger; 1983.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington (DC): American Psychiatric Press; 1994.
3. Jacobs DM, Sano M, Dooneief G, Marder K, Bell KL, Stern Y. Neuropsychological detection and characterization of preclinical Alzheimer's disease. Neurology 1995;45:957-962.
4. Knesevich JW, LaBarge E, Edwards D. Predictive value of the Boston Naming Test in mild senile dementia of the Alzheimer type. Psychiatry Res 1986;19:155-161.
5. Kittner SJ, White LR, Farmer ME, Wolz M, Kaplan E, Moes E, et al. Methodological issues in screening for dementia: the problem of education adjustment. J Chronic Dis 1986;39:163-170.
6. Kim H, Na D. Korean Version-Boston Naming Test. Seoul: Hakjisa; 1997.
7. Kim HH, Na DL. Normative data on the Korean version of the Boston Naming Test. J Clin Exp Neuropsychol 1999;21:127-133.
8. Kang YW, Kim HH, Na DL. A short form of the Korean-boston naming test (K-BNT) for using in dementia patients. Korean J Clinl Psychol 1999;18:125-138.
9. Kang YW, Kim HH, Na DL. Parallel short forms for the Korean-Boston naming test (K-BNT). J Korean Neurol Assoc 2000;18:144-150.
10. Lee JH, Lee KU, Lee DY, Kim KW, Jhoo JH, Kim JH, et al. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. J Gerontol B Psychol Sci Soc Sci 2002;57:47-53.
11. Pedraza O, Graff-Radford NR, Smith GE, Ivnik RJ, Willis FB, Petersen RC, et al. Differential item functioning of the Boston Naming Test in cognitively normal African American and Caucasian older adults. J Int Neuropsychol Soc 2009;15:758-768.
12. Ramirez M, Teresi JA, Holmes D, Gurland B, Lantigua R. Differential item functioning (DIF) and the Mini-Mental State Examination (MMSE). Overview, sample, and issues of translation. Med Care 2006; 44(11 Suppl 3):S95-S106.
13. Jones RN, Gallo JJ. Education and sex differences in the mini-mental state examination: effects of differential item functioning. J Gerontol B Psychol Sci Soc Sci 2002;57:548-558.
14. Teresi JA, Golden RR, Cross P, Gurland B, Kleinman M, Wilder D. Item bias in cognitive screening measures: comparisons of elderly white, Afro-American, Hispanic and high and low education subgroups. J Clin Epidemiol 1995;48:473-483.
15. Marshall S, Mungas D, Weldon M, Reed B, Haan M. Differential item functioning in the Mini-Mental State Examination in English-and Spanish-speaking older adults. Psychol Aging 1997;12:718-725.
16. Teresi J, Kleinman M, Ocepek-Welikson K, Ramirez M, Gurland B, Lantigua R, et al. Applications of item response theory to the examination of the psychometric properties and differential item functioning of the comprehensive assessment and referral evaluation dementia diagnostic scale among samples of Latino, African American, and White non-Latino elderly. Res Aging 2000;22:738-773.
17. Weinstock LM, Strong D, Uebelacker LA, Miller IW. Differential item functioning of DSM-IV depressive symptoms in individuals with a history of mania versus those without: an item response theory analysis. Bipolar Disord 2009;11:289-297.
18. Kim KW, Park JH, Kim MH, Kim MD, Kim BJ, Kim SK, et al. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. J Alzheimers Dis 2011;23:281-291.
19. Lee DY, Lee KU, Lee JH, Kim KW, Jhoo JH, Kim SY, et al. A normative study of the CERAD neuropsychological assessment battery in the Korean elderly. J Int Neuropsychol Soc 2004;10:72-81.
20. Reise SP, Ainsworth AT, Haviland MG. Item response theory fundamentals, applications, and promise in psychological research. Curr Dir Psychol Sci 2005;14:95-101.
21. McGrory S, Doherty J, Austin E, Starr J, Shenkin S. Item response theory analysis of cognitive tests in people with dementia: a systematic review. BMC Psychiatry 2014;14:47.
22. Teresi JA. Mini-Mental State Examination (MMSE): scaling the MMSE using item response theory (IRT). J Clin Epidemiol 2007;60:256-259.
23. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002;27:77-83.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B Met 1995;57:289-300.
25. Crane PK, Gibbons LE, Jolley L, van Belle G, Selleri R, Dalmonte E, et al. Differential item functioning related to education and age in the Italian version of the Mini-mental State Examination. Int Psychogeriatr 2006;18:505-515.
26. Holman R, Glas CA, de Haan RJ. Power analysis in randomized clinical trials based on item response theory. Control Clin Trials 2003;24:390-410.
27. Zec RF, Burkett NR, Markwell SJ, Larsen DL. Normative data stratified for age, education, and gender on the Boston Naming Test. Clin Neuropsychol 2007;21:617-637.
28. Nicholas M, Obler L, Albert M, Goodglass H. Lexical retrieval in healthy aging. Cortex 1985;21:595-606.
29. Ripich DN, Petrill SA, Whitehouse PJ, Ziol EW. Gender differences in language of AD patients: a longitudinal study. Neurology 1995;45:299-302.
30. Buckwalter JG, Sobel E, Dunn ME, Diz MM, Henderson VW. Gender differences on a brief measure of cognitive functioning in Alzheimer's disease. Arch Neurol 1993;50:757-760.
31. Bayles K, Azuma T, Cruz R, Tomoeda C, Wood J, Montgomery Jr E. Gender differences in language of Alzheimer disease patients revisited. Alzheimer Dis Assoc Disord 1999;13:138-146.
32. Kurt M, Karakaya İ, Safaz İ Ateş G. Differential item functioning by education and sex in subtests of the Repeatable Battery Assessment of Neuropsychological Status. Eur J Psychol Assess 2015;31:5-11.
33. Crane P, Gibbons L, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIF detect and difwithpar. Med Care 2006;44(11 Suppl 3):S115-S123.
34. Lukatela K, Malloy P, Jenkins M, Cohen R. The naming deficit in early Alzheimer's and vascular dementia. Neuropsychology 1998;12:565-572.
35. Diehl J, Monsch AU, Aebi C, Wagenpfeil S, Krapp S, Grimmer T, et al. Frontotemporal dementia, semantic dementia, and Alzheimer's disease: the contribution of standard neuropsychological tests to differential diagnosis. J Geriatr Psychiatry Neurol 2005;18:39-44.