Educational Sciences: Theory & Practice - 12(1) • Winter • 382-385

www.edam.com.tr/estp

Assessment of Dimensionality in Social Science Subtest

Özlem Yeşim ÖZBEK BAŞTUĞ[a]
Çankırı Karatekin University

Abstract

Most of the literature on dimensionsionality focused on either comparison of parametric and nonparametric dimensionality detection procedures or showing the effectiveness of one type of procedure. There is no known study to shown how to do combined parametric and nonparametric dimensionality analysis on real data. The current study is aimed to fill this missing part in the literature by illustrating how to do combined parametric and nonparametric dimensionality analysis. The purpose of this study is to describe dimensionality structure of social science subtest of the Secondary School Institutions Student Selection and Placement Test using combined parametric and nonparametric dimensionality analysis. The data from the social science subtests of the Secondary School Institutions Student Selection and Placement Test of 1999, 2000, and 2001 were used for this study. The study indicated multidimensionality for the social science subtest. Because the results indicated multidimensionality does exist in social science subtest, it would be helpful to describe multidimensionality structure and, finally, score separately by these unidimensional grouping.

Key Words

Dimensionality, Unidimensionality, Validity, Parametric Methods, Nonparametric Methods, Social Science Subtest.

Claiming unidimensionality does not itself ensure the validity of the test, and any assumption of unidimensionality should be checked. Unidimensionality defined as the existence of one latent trait or construct underlying a set of measures (Hattie, 1985; McDonald, 1981). Procedures used to assess the dimensionality are profound and based on various techniques. However, most of the literature focused on either comparison of parametric and nonparametric dimensionality detection procedures (e.g., Finch & Habing, 2003; Mroch & Bolt, 2006) or showing the effectiveness of one type of procedure (e.g., Roussos, Stout, & Marden, 1998; Stout, Froelich, & Gao, 2001). Only a few studies have illustrated how to do dimensionality analysis either using parametric or nonparametric methods on real data (e.g., Douglas, Kim, Roussos, Stout, & Zhang, 1999; Jang & Roussos, 2007). Also, no known study has shown to how to do combined parametric and nonparametric dimensionality analysis on real data.

Therefore, current study is aimed to fill this missing part in the literature by illustrating how to do combined parametric and nonparametric dimensionality analysis. The second purpose of this study is to describe dimensionality structure of social science subtest of the Secondary School Institutions Student Selection and Placement Test (SSISSPT). Describing the dimensional structure (e.g., verification of unidimensionality or multidimensionality) is important to confirm the construct equivalence of social science subtest across gender and forms. Furthermore, verification of

a **Özlem Yeşim ÖZBEK BAŞTUĞ, Ph.D.**, is currently an Assistant Professor at the Department of Education. Her research interests include validity, dimensionality, differential item functioning. Correspondence: Assistant. Prof. Dr. Özlem Yeşim Özbek Baştuğ, Çankırı Karatekin University Faculty of Literature, Department of Education, 18100 Çankırı / Turkey. E-mail: ozbek@karatekin.edu.tr. Phone: +90 376 2541215/4584.

unidimensionality is important because many IRT techniques (e.g., BILOG) presume unidimensionality of the data. Use of these IRT procedures can be justified by a statistical analysis to confirm approximate unidimensionality or by statistical argument to claim that the departure from unidimensionality is not serious enough to jeopardize use of specific tools (Stout, 1987).

Method

Data

The data from the social science subtests of the Secondary School Institutions Student Selection and Placement Test of 1999, 2000, and 2001 in Turkey were used for this study. Each year's data contains responses from approximately 350,000 examinees. Two random samples of 4000 examinees were drawn from each data set with equal number of female and male examinees to perform dimensionality analyses.

The social science subtest was constructed to measure students' general social science knowledge (e.g., remembering the particular knowledge on history, geography), social science conception and notion knowledge (e.g., being able to interpret graphs and maps, providing examples, transforming particular knowledge), application skills (e.g., finding required principles and rules, applying a generalization, reading a map or graph to solve a problem), and social analysis skill (e.g., being able to separate general knowledge to its parts, finding the relationship, principles, rules, deficiencies or harmonies in given general knowledge). An approximate percentage of items in the social science subtest could be given as: 41 %, 27 %, 16 %, and 16 %, respectively, for history and Turkish history items, geography items, citizenship items, and religion and ethic items.

Statistical Procedures

Descriptive statistic analysis (means, standard deviations, variances, reliability indices, kurtosis, and skewness) and dimensionality analysis were performed on the data. Assessment of test unidimensionality is generally performed at two levels: firstly, the assumption of unidimensionality is checked; if the hypothesis of unidimensionality is rejected, then detailed investigation is performed to describe underlying multidimensional structure. Statistical and nonstatistical dimensionality assessment methods were used. Specifically, combination of several techniques applied in this study included the use of several non-parametric dimensionality assessment tools (DIMTEST, ATFIND, DETECT, HCA/CCPROX), subjective dimensionality analyses, a nonlinear factor analytic dimensionality analysis tool (NOHARM) and a unidimensional IRT item estimation tool (BILOG). Each step of the study revealed unique and important information relevant to dimensionality investigation. The general approach applied to explore dimensionality structure in social science subtests was as follows: (*a*) BILOG program was run on each subtest to obtain item statistics and parameter estimates; (*b*) blind to statistical analysis an extensive content analysis was performed on each social science subtest; (*c*) exploratory DIMTEST and DETECT analyses were performed on each social science subtest; (*d*) if the results of exploratory DIMTEST and DETECT indicated unidimensionality for a given subtest, analyses were stopped and no further action was taken on the subtest; (*e*) however, when the results of exploratory DIMTEST and DETECT indicated multidimensionality for a specific subtest, several follow up analyses were conducted to explore underlying dimensionality structure. Because content analysis indicated multidimensionality could be due to differences in the item content area, follow up analysis began with a confirmatory DIMTEST and DETECT analysis to test whether the content based clustering was the underlying reason for presented multidimensionality. To further investigate the sources of multidimensionality a sequential DIMTEST-HCA/CCPROX analysis was conducted. The results of exploratory DETECT analysis were also useful to understand sources of dimensionality, so the sign matrices and clusters from the exploratory DETECT analysis examined extensively. Semi-exploratory NOHARM analysis was originally planned to be a part of dimensionality analysis for each subtest to clarify the structure if the test indicated an approximate simple structure; however, NOHARM indicated serious estimation problems, presumably due to small item number used in subtests (NI=25). Therefore, NOHARM results were not reported at the subtest level. Finally, hypotheses on the dimensional structure of the SSISSPT were developed based on what had been learned from dimensionality analysis.

Results

The summary descriptive statistics and reliability estimates present males slightly outperformed females across all three administrations. The reliabil-

ity indices for the social science subtest ranged between .74 and .83. All values for variable skewness and kurtosis were within the limits.

The item content analysis of the social science subtest yielded four categories of distinct skills in each form: geometry, history, citizenship and selective questions. The social science subtest includes a selective section, which consists of either four religion or social science items (history and geography) based on a student's preferred area.

Exploratory DETECT analysis indicated social science subtests for 1999, 2000 and 2001 were multidimensional. Supporting DETECT, exploratory DIMTEST analysis also rejected unidimensionality for the social science section of SSISSPT for all three years with very small p-values.

In confirmatory DIMTEST analysis, DIMTEST rejected the hypotheses of unidimensional similarity between each content area and the remainder of the subtest for all 12 runs. Thus, it appears that each content area introduced its own distinct dimension into the social science subtest. Next, the amount of multidimensionality was addressed by confirmatory DETECT. Being very close to results produced by exploratory DETECT, DETECT indexes were between 0.24 and 0.30 for the confirmatory analysis. Supporting the hypothesis that the multidimensionality of social science subtest was associated with the item content area, sign matrices provided by confirmatory DETECT analysis showed three clear clusters.

Beginning at the early stages of the analysis, HCA indicated three major clusters: a combined cluster of citizenship and religion items, a cluster of geography items and a cluster of history items for both 2000 and 2001. Analysis also indicated a cluster of two bad items, which finally joined the rest of the test at the last stage of the analysis in 2000 and 2001. The HCA results were more complicated for the 1999 administration. At stage eighteen, there were two big clusters and several individually clustered items. One of the clusters included mostly history items and the other one was formed by citizenship and three geography items. In the later stages of the analysis, the big clusters and most of the individually clustered items joined together to form one big cluster. Only at final stage a three-item cluster, which again could be described as bad items due to their low item test correlation, joined the rest of the social science items.

The results of sequential DIMTEST-HCA analysis for social science subtest for the three data sets presents the number of dimensionally distinct clusters was three for all three forms. AT sets were chosen as in the previous subtests. Close investigation of these clusters showed that except 1999 each cluster closely matched to a social science content area.

Exploratory DETECT analysis immediately shows there were only three dimensions in social science subtest. A close investigation of these clusters showed three clusters were matched exactly to three content areas: history, geography and combination of citizenship and selective questions. Although the percentage of students who answered the religion questions was not known, by looking at the clusters produced by DETECT one could argue most of the students preferred to answer religion questions rather than social science questions because none of the selective questions clustered with history or geography items across three administrations. There were a few puzzling items in each form. In 1999, items 4, 5 and 14 did not joined to big clusters, however, after close inspection of these items one could argue that these items did not require any specific social science knowledge and could be answered by only reading comprehension. In 2000, two history items (11, 12) clustered with the geography items and items 9 and 21 were not clustered with their own content area. Close inspection of these items showed items 11 and 12 actually required geography knowledge and items 9 and 21 very much resembled reading comprehension questions. In 2001, items 1, 8 and 20 were not clustered with their expected groups. A close investigation of these items revealed items 1, 8 and 20 had very low item-test correlations and again could be categorized as reading comprehension questions rather than social science questions.

Discussion

The study indicated multidimensionality for the social science subtest. Because the results indicated multidimensionality does exist in social science subtest, it would be helpful to describe multidimensionality structure in subtest, divide each scale according to its underlying dimensionality structure and, finally, score separately by these unidimensional grouping. This approach might prevent several multidimensionality based scoring problems and allow application of unidimensionality-based procedures.

There were several important findings for the social science subtest. First, content analysis indicated items in social science subtest came from four major content areas: geometry, history, citizenship and a selective area which either included religion items

or a combination of history, geometry and citizenship items. When item content based clusters were tested using DIMTEST and DETECT to see if in fact difference in item content area was the cause of detected multidimensionality in the social science subtest, findings confirmed that the content difference was an important part of the multidimensional structure. This finding was also supported by the HCA/CCPROX analysis for 2000 and 2001 administrations. The results for 1999 were not as clear as these for subsequent administrations, possibly because of a few bad items with low item test correlations in the 1999 test. However, it is also possible that the test structure of the 1999 administration might be considerably different from that of the 2000 and 2001 administrations; in fact, substantive analysis did indicate some evidence for this. Findings from HCA/CCPROX were also supported by sequential use of DIMTEST and HCA/CCPROX. In line with the previous findings, exploratory DETECT analysis indicated item content differences in social science subtest were responsible for observed multidimensionality. Another important finding was related to NOHARM analysis. Unfortunately, the NOHARM procedure which was proven to be a very useful method for finding test structure during dimensionality investigation of the entire test, failed to provide dependable solutions for the subtests. However, subtests had only one-fourth of entire SSISSPT items, and this could be the reason for NOHARM convergence problems. To overcome this problem, maybe items from the different forms could be combined and the analysis repeated on these bigger item sets in the future.

Increasing number of items in each subtest is recommended to reliably estimate student ability or achievement in that particular area. The current SSISSPT has only 25 items for each subtest. Also, keeping the same number of items for primary content areas as well as for secondary content areas across different administrations is recommended for SSISSPT. Currently, forms do show several inconsistencies and this could cause several inequity issues across test-takers. Limited numbers of items in each subtest may have prevented some of the procedures from producing reliable results. Therefore, combining the subtests across different administrations, the dimensionality structure of the data can be further explored. It should be underlined that this current study is only a preliminary step in revealing the dimensionality structure of SSISSPT and certainly an incomplete one. Thus, more in depth analysis for each subtest is required to fully understand the specific structure of subtests.

References/Kaynakça

Douglas, J., Kim, H. R., Roussos, L., Stout, W., & Zhang, J. (1999). *LSAT Dimensionality analysis for the December 1991, June 1992, and October 1992 Administrations* [Law School Admission Council Statistical Report 95-05].

Finch, H., & Habing, B. (2003, April). *Comparison of NOHARM and DETECT in item cluster recovery: Counting dimensions and allocating items.* Paper presented at the annual meeting of the National Council on Measurement, Chicago.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of testes and items. *Applied Psychological Measurement, 9*, 139-164.

Jang, E. E., & Roussos, L. A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based non-parametric approach. *Journal of Educational Measurement, 44* (1), 1-21.

McDonald, R. P. (1981). The dimensionality of test and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117.

Mroch, A. A., & Bolt, D. M. (2006). A simulation comparison of parametric and nonparametric dimensionality detection procedures. *Applied Measurement in Education, 19* (1), 67-91.

Roussos, L. A., Stout, W. F., & Marden, J. L. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, *35*, 1-30.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.

Stout, W., Froelich, A. G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essay on item response theory* (pp. 357-375). New York: Springer.