CONFIRMATORY AND EXPLORATORY FACTOR ANALYSES OF STUDENTS' DEVELOPMENTAL LEVELS IN LEARNING STATISTICS

Rini Oktavia[1] and M. Alejandra Sorto[2]
[1]Universitas Syiah Kuala, Indonesia & Valparaiso University, USA
[2]Texas State University, USA
rini.oktavia@valpo.edu

This study provides an empirical based-explanation on how students develop their understanding of statistical concepts and investigations. We developed an instrument that measures students' developmental levels in learning statistics and administered it to 797 high school and middle school students in central Texas. We conducted exploratory and confirmatory factor analyses by applying structural equation modeling (SEM) approaches. In this paper, we discuss the results of these analyses.

INTRODUCTION

Two important US Pre-K-12 curriculum documents for statistics are the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework (Franklin, et al., 2007) and the Common Core State Standards in Mathematics (CCSS-M) (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). These documents provide detailed descriptions of what students should know and should be able to do in order to develop sound statistical reasoning. The Pre-K-12 GAISE Framework presents statistical problem solving into four process components: (1) formulating questions, (2) collecting data, (3) analyzing data, and (4) interpreting results. The framework also explains the developmental process on students' understanding of the nature of variability and the focus of variability. The framework hypothesizes that students develop their understanding on each statistical process component through three developmental levels (Levels A, B, and C). Although the basic structure of the process is the same for all levels, the degree of sophistication of statistical problems addressed over the developmental levels increases (Peck, Kader, & Franklin, 2008). Even though this progression is well described, it is based partially on theory and has not yet been empirically tested. Therefore, in this study, we provide an empirical-based description of how students develop their statistical knowledge and skill.

METHODOLOGY

To accomplish the goal of this study, we developed two instruments: the expert survey instrument and the student survey instrument. We developed 40 items for the student instrument in order to measure students' developmental levels and learning trajectories in statistics. The expert survey asked experts to align the items in the student instrument with their GAISE levels and to validate the statistical content addressed by the items. The items were adapted from the Statistical Reasoning Assessment (Garfield J. 1991, 2003), the Statistical Literacy Assessment (Callingham & Watson, 2005;Watson J. M., 1997; Watson & Callingham, 2003), the Assessment Research Tool for Improving Statistical Thinking (ARTIST) Project database, and released items from the Program for International Student Assessment (PISA) (OECD, 2009). We also developed new items based on the Pre-K-12 GAISE Framework and the CCSS-M guidelines. The items include six major ideas in probability and statistics: (1) awareness of statistical question distinction, (2) sampling and data collection methods, (3) measures of centers (averages), (4) variability, (5) graph representation and interpretations, and (6) association, covariation, and correlation. We organized the six ideas into the four statistical investigation process components. Some of the items also assess students' understanding of the nature of variability and the focus on variability. Two examples of the items in the student instrument that investigate students' understanding on simple sampling methods and on the measures of center (averages) are presented in Example 1 and Example 2 below.

In the summer of 2012, we conducted a pilot study with 85 participants. The results of item analysis in the pilot study suggested that only 36 of the 40 items in the pilot instrument have acceptable psychometric qualities. We also conducted an expert survey that asked experts to

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

analyze the statistical contents involved in the items of the pilot instruments and to align them to the GAISE levels. The results of item analyses and the expert survey during pilot study showed that among the 36 acceptable items, 10 items were aligned to Level A, 16 items were aligned to Level B, and 10 items were aligned to Level C. The pilot study also revealed that students, on average, took about two to three minutes to answer an item in the instrument.

Example 1. (ITEM 04 FORM 1)
A farmer wants to know how many fish there are in his dam. He took out 200 fish and tagged each of them with a colored sign. He put the tagged fish back in the dam and let them get mixed with the others. On the second day, he took out 250 fish randomly and found that 25 of them were tagged. Estimate how many fish are in the dam.
A. 250 *B. 500* *C. 1000* ***D. 2000***
(Adapted from Watson & Callingham, 2003).

Example 2. (ITEM 30 FORM 2)
A town contains three elementary schools. School A has a mean class size of 30 pupils for its three fifth-grade classrooms. School B has a mean class size of 25 pupils in its two fifth-grade classrooms. School C has 20 pupils in its only fifth-grade classroom. What is the average class size for fifth-grade classrooms in this town?
A. 12.5 *B. 25* ***C. 26.7*** *D. Cannot be determined.*

For practical purposes, we distributed the 36 items between two forms, each including a similar proportion of items from each level and each content area. With this distribution, we predicted that it would take less than 60 minutes to administer the survey. This prediction was proved correct when we administered the survey to 797 students ranging from Grade 6 to Grade 12 in the fall of 2012. Among the 797 participants, 140 participants received Form 1 and 657 students received Form 2. We also conducted another expert survey that asked similar questions to those that we asked in the pilot study.

Following the experts' opinions, we developed five structural equation models that describe the alignment between the items and the levels for each form. The models also describe the hierarchical relationship among the GAISE levels. We developed the first model based on experts' opinion during the pilot study. We also developed the second and third models following the opinions of two experts (Expert 1 and Expert 2). We conducted an exploratory factor analysis using a structural equation modeling (SEM) approach to find the best model for each form. Then, the model found became the fourth model called the combination model. Finally, we developed the fifth model, the reduced combination model, by reducing low performing items from the combination model and applying an SEM-based confirmatory factor analysis (CFA).

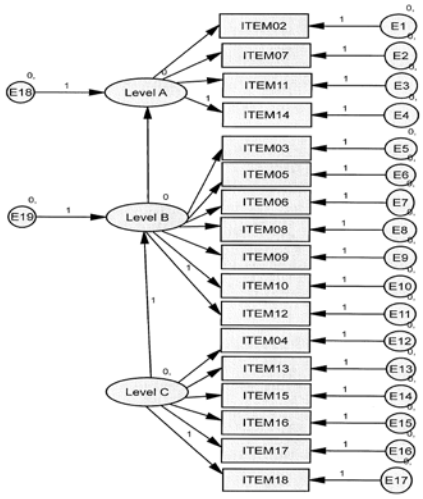

Figure 1 Initial Model (Form 1)

Due to limited space, only the SEM diagram for the initial model based on experts' survey during the pilot study model is presented in this paper (see Figure 1). Similar diagrams were created for the other models. The model delineates that Level C mastery impacts Level B mastery and Level B mastery impacts Level A mastery as suggested by the Pre-K- 12 GAISE Framework. The model consists of three unobserved latent variables: Level A, Level B, and Level C, seventeen observed variables (ITEM02 – ITEM18) and 19 residual errors associated with the observed variables (E1-E19). The model shows that ITEMs 02, 07, 11, and 14 are aligned to Level A; ITEMs 03, 05, 06, 08, 09, 10, and 12 are aligned to Level B; ITEMs 04, 13, 15, 16, 17, and 18 are aligned to Level C.

RESULTS

From regression and standardized regression weight parameter estimates of the first four models of Form 1, we found that the relation between the latent variables Level B and Level C as well as Level A and Level B are significant (p-value < 0.001 levels, 2 tailed). All standardized factor loadings of the latent variables are larger than 0.6. This indicates that the theory suggesting students develop their understanding starting from Level A, then moving to Level B, and finally reaching Level C is confirmed by all four models. Likewise, all models for Form 2 show that the relation between the latent variables Level B and Level C as well as between Level A and Level B are significant (p-value < 0.001 levels, 2 tailed). After analyzing the regression weight estimates, it was found that several items are problematic. By reducing problematic items in the combination models of both forms, we developed the Reduced Combination F1 model and the Reduced Combination F2 model. From regression and standardized regression weight parameter estimates of both models, it is revealed that all parameter estimates of each item in both models are reasonable and statistically significant at 0.05 levels for Form 1 and at 0.001 levels for Form 2. The standard errors (S.E.) and the critical ratios (C.R.) of all items in these two forms are also in good order (C.R. > 1.96).

For further use of the instrument, it is essential to confirm which model best fits the data. An SEM-based model fit was analyzed using the MPlus Version 7 program. The standard fit statistics used are the minimum discrepancy (CMIN), the degree of freedom (DF), the CMIN/DF ratio, the p-value, the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), and the root mean square error of approximation (RMSEA). The CMIN/DF ratios less than 2.0 are indicators of a good fit, and ratios greater than 2.0 but less than 3.0 are indicators of a modest fit (Purpura & Lonigan, 2013). Large p-values indicate that the data is consistent with the hypothesized model. CFI and TLI values of greater than .95, and RMSEA values of less than or equal to .05 are indicators of a good fit (Purpura & Lonigan, 2013). In addition, Hu and Bentler (1999) recommend that an RMSEA value between .05 and .08 is an indicator of a moderate fit.

The results showed that all five models of Form 1 fit the data well (CMIN/DF < 2.00; CFI > .95; TLI > .95; and RMSEA < .05). The results also revealed that the Initial Model of Form 2 does not fit the data well (p-value < 0.05; CFI < .95; and TLI < .95). Likewise, the Expert 1's Model of Form 2 does not fit the data well (p-value ≤ 0.05, CFI < .95, and TLI < .95). On the other hand, the Expert 2's Model and the Combination Model of Form 2 fit the data well (CMIN/DF < 2.00; CFI > .95; TLI > .95; and RMSEA < .05). For the Reduced Combination F2 Model, it is found that the p-value slightly meets the criteria of good fit models for .01 significant levels (p-value > 0.01). All other criteria are fulfilled satisfactorily (CMIN/DF < 2.00; CFI > .95; TLI > .95; and RMSEA < .05). Therefore, we can conclude that the reduced combination models fit the data well. Detailed results are presented in the author's dissertation (Oktavia, 2013).

CONCLUSION

From an SEM-based CFA of both forms, it was found that all models for Form 1 and three of the five models of Form 2 fit the data well. After analyzing the regression weight estimates, it was found that several items are problematic. The new models developed by removing the problematic items in the combination model of both forms suggest that the removed items should be excluded from the instrument. This suggestion, however, could result in a shorter length of the survey form that might compromise the internal consistency reliability of items for each form.

It is convincing that the items included in the Reduced Combination Models of Form 2 are suitable to measure students' developmental level as suggested by the Pre-K-12 GAISE Framework. Because it has been administered to a larger sample and after reducing its problematic items, Form 2 will be the best form to identify students' developmental level in statistics. The instrument is available for use and can be requested by contacting the authors.

Finally and most importantly, the results showed that the hypothesis that students develop their understanding by starting with Level A, then working successfully through Level B, and finally through Level C, as suggested by the Pre-K-12 GAISE Framework, is supported by the data. These results provide an empirical-based description on how students develop their statistical knowledge and skill that agrees with the hypothetical description presented by the Framework. However, the instrument developed in this study does not cover all statistical concepts that high school and middle school students should be able to do, as suggested by the Pre-K-12 GAISE Framework and the CCSS-M. Claiming that the instrument can precisely diagnose students' developmental level and learning trajectories in statistics would be misleading. Further studies must be conducted to develop a more rigorous instrument to measure students' developmental levels and learning trajectories in statistics.

ACKNOWLEDGEMENTS

This study was funded by the Graduate College of Texas State University through the Dissertation Research Stipend Award and by the Department of Mathematics at Texas State University.

REFERENCES

Callingham, R. A., & Watson, J. (2005). Measuring statistical literacy. *Journal of Applied Measurement, 6*(1), 19-47.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework.* Alexandria, VA: American Statistical Association.

Garfield, J. (1991). Evaluating students' understanding of statistics: Development of the Statistical Reasoning Assessment. *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Volume 2, pp. 1-7). Blacksburg, VA: VPI and SU.

Garfield, J. (2003). Assessing statistical reasoning. *Statistical Education Research Journal, 2*(1), 22-38.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards-Mathematics.* Washington D. C. : National Governors Association Center for Best Practices, Council of Chief State School Officers.

Oktavia, R. (2013). *Identifying developmental level and learning trajectories in statistics for grade 6-12 students.* Dissertation. Texas State University-San Marcos.

Organization for Economic Cooperation and Development (OECD) (2009). *Take the Test: Sample Questions from OECD's PISA Assessment*. Paris: Author.

Purpura, D. J., & Lonigan, C. J. (2013). Informal numeracy skills: The structure and relations among numbering, relations, and arithmetic operations skills in preschool. *American Education Research Journal, 50*, 178-209.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. al, & J. B. Garfield, *The assessment challenge in statistics education* (pp. 107–121). Amsterdam, The Netherlands: IOS Press and The International Statistical Institute.

Watson, J. M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *StatisticsEducation Research Journal, 2*(2), 3-46.