The Physical Educator Vol. 72 • pp. 206–223 • 2015

ASSESSMENT

Validity and Responsiveness of Concept Map Assessment Scores in Physical Education

Yun Soo Lee, Yongkyu Jang, Minsoo Kang

Abstract

Concept map assessment has been applied to many education areas to measure students' knowledge structure. However, the proper and valid use of concept map assessment has not been examined in physical education. The purpose of this study was to evaluate the evidence of validity and responsiveness of the concept map assessment scores in physical education teacher education (PETE) programs. Concept map data were collected from 56 students. Two raters independently scored concept maps using structural method and relational method. Intraclass correlation coefficient (ICC) was used to examine interrater reliability. Independent t tests and paired t tests were used to examine validity and responsiveness. High level of reliability was seen between two raters (ICC = .946–.985). The results of this study provided the evidence of validity and responsiveness. Concept map assessment reflects expected difference between the groups and the change in students' knowledge structure over time. In conclusion, the feasibility of the concept map assessment in PETE was identified in this study.

Yun Soo Lee is an assistant professor, Department of Physical Education, Dankook University. Yongkyu Jang is a professor, Department of Physical Education, Seoul National University of Education. Minsoo Kang is a professor, Department of Health and Human Performance, Middle Tennessee State University. Please send author correspondence to Yun Soo Lee, Department of Physical Education, College of Education, Dankook University, Yongin, South Korea. E-mail: yslee@dankook.ac.kr

Providing high-quality education and training to teacher candidates in teacher education programs is important. In 2008, for example, the National Association for Sport and Physical Education (NASPE) proposed national standards for initial physical education teacher education (PETE) to establish high-quality PETE programs in the United States. Teacher candidates in PETE programs need to be provided with acceptable knowledge, skills, and dispositions (National Council for Accreditation of Teacher Education [NCATE], 2008). PETE programs also need to be evaluated using the standards (e.g., NASPE, NCATE) that were adopted for the program. Consequently, it becomes necessary to assess teacher candidates' knowledge and skills using valid and reliable assessment tools.

Traditionally, teacher educators have used a variety of assessments to measure teacher candidates' knowledge and skills. One of the most widely used assessment tools is a written test (Lund & Kirk, 2002), which comprises short-answer questions, multiple-choice questions, essay questions, matching questions, true–false questions, classification questions, and rearrangement questions. Journals, lesson planning, projects, and portfolios are also often used as assessment tools in teacher education programs (Lund & Kirk, 2002). However, these traditional assessments have limitations for assessing teacher candidates' knowledge. First, these tools are used to assess only limited and lower order knowledge (e.g., remembering and comprehension). Second, it takes much time to assess and grade teacher candidates' in-depth knowledge. For these reasons, teacher educators are in need of new and better ways to assess teacher candidates' knowledge.

Concept map assessment is an alternative way of assessing teacher candidates' knowledge structure. A concept map is a diagram "to represent meaningful relationships between concepts in the form of propositions" (Novak & Gowin, 1984, p. 15). Concept maps include main ideas (i.e., key concepts) that are enclosed in circles with arrows to connect one concept (i.e., one circle) to another. On the arrow lines, there are linking words or phrases to explain the relationship between the two concepts. Propositions are developed in concept maps, including two or more concepts with linking words or phrases to make a meaningful statement (Novak & Cañas, 2008). Joseph Novak developed concept maps in his research program at Cornell in 1972 (Novak & Cañas, 2008). This research program was based on David Ausubel's (1968) cognitive psychology theory (as cited in Novak & Cañas, 2008), of which the principle concept is that “learning takes place by the assimilation of new concepts and propositions into existing concepts and propositional frameworks held by the learner” (Novak & Cañas, 2008, p. 3).

Concept map assessment is an effective tool to demonstrate an individual’s organization of knowledge and decision making (West, Pomeroy, Park, Gerstenberger, & Sandoval, 2000). It may also be used to measure teacher candidates’ abilities to apply, analyze, synthesize, and evaluate. It is simple and easy to teach teacher candidates to generate concept maps (McClure, Sonak, & Suen, 1999; Rice, Ryan, & Samson, 1998) and requires little time for teacher educators to grade them. An individual’s knowledge structure can be expanded with new knowledge plus existing knowledge (Novak & Cañas, 2008). As a result of the connections between new and existing knowledge, teacher candidates can gain a deeper understanding of the topic. Furthermore, the process can be repeated to show evidence of how teacher candidates are integrating new knowledge into existing concept maps.

Research on concept maps has been developed in science education, wherein researchers measure students’ knowledge by scoring the structure of the concept maps or the relationships among the concepts. McClure et al. (1999) established reliability, validity, and logistical practicality for concept map assessment using three scoring methods by comparing the students’ concept maps with a master map. They suggested that concept maps can be used to find unique and valuable information about students’ knowledge structure (McClure et al., 1999). Other researchers have tried to measure procedural knowledge as well as declarative knowledge using concept maps (Rice et al., 1998).

Shulman (1986) proposed his view of the knowledge for teaching (a) subject matter content knowledge (CK), (b) pedagogical content knowledge (PCK), and (c) curricular knowledge. Teacher educators in educational communities are interested in teacher knowledge and attempt to develop and assess teacher candidates’ knowledge effectively. Therefore, it is important to assess teacher candidates’ different knowledge accurately (e.g., CK, PCK, curricular knowledge) using different assessment tools. If teacher educators use concept maps as an alternative assessment tool, it would be beneficial to measure teacher candidates’ knowledge base for teaching that which Shulman described. The reason is because it takes little time for teacher educators to train teacher candidates and to score concept maps. It is also not difficult for teacher candidates to create concept maps in terms of time and skills.

Concept map assessment has been recommended for use in subject matters such as biology education (Pearsall, Skipper, & Mintzes, 1997), engineering education (Besterfield-Sacre, Gerchak, Lyons, Shuman, & Wolfe, 2004), science education (McClure & Bell, 1990; McClure et al., 1999; Rice et al., 1998; Ruiz-Primo & Shavelson, 1996; Rye & Rubba, 2002), medical education (Kassab & Hussain, 2010; West, Park, Pomeroy, & Sandoval, 2002; West et al., 2000), physics education (Austin & Shore, 1995), and physical education (PE; Ennis, Mueller, & Zhu, 1991; Mohammed, 2010; Rink, French, Lee, Solomon, & Lynn, 1994). However, the proper scoring methods of a concept map and the validity and responsiveness of concept map assessment have not been examined in PETE programs. The purpose of this study was to evaluate the evidence of validity and responsiveness of concept map assessment scores in PETE programs.

Method

Participants

The participants were 57 students (i.e., 31 students in Group A and 26 students in Group B) who were recruited from the PETE program at a university in the southeast region of the United States. A power analysis was performed prior to this research study based on West et al.'s (2000) data. Their concept map data were used, and test of difference between two independent means was selected to estimate sample size of this study. The results indicate that a sample size of 52 (26 in each group) is necessary, given the effect size of 0.92, alpha of .05, and the power of 0.90.

This study was approved by the institutional review board at this university. A written informed consent form was collected from each participant. The participants in Group A were students who were taking a course, Assessment in Physical Education, which is a senior-level course according to the sequence of the program. The participants in Group B were students who were taking Introduction for Teaching Physical Education, a sophomore-level course and an initial course for PE-related majors (e.g., health education, leisure and recreation, exercise science). Students in Group A were requested to complete two concept maps, and students in Group B were requested to complete one concept map. One student in Group A did not complete a second concept map, resulting in removal from further analysis. Therefore, 56 participants' data (i.e., 30 students in Group A and 26 students in Group B) were used in this study.

Setting

The participants completed concept map training and created a concept map about the topic, assessment in PE. This topic was selected because (a) the use of quality assessment is one of the important concepts that teacher candidates should obtain from a PETE program; (b) this topic is aligned with National Standard 5, impact on student learning (NASPE, 2008); and (c) the researcher is familiar with this topic to score concept maps accurately as a rater.

Instructions about assessment in PE were provided to only Group A using the course textbook (Lund & Veal, 2013). Using this textbook, students learned about key concepts related to assessment with a number of related activities to consolidate student understanding. The instructor taught key concepts of the assessment in PE. The participants in Group A read and completed all tasks in Chapters 1 to 14. The instructor provided examples about key concepts and showed how to develop assessment plans for PE classes. The participants had opportunities to discuss the concepts within small groups and as a whole group. In addition, the participants had opportunities to use and experience different assessments while engaged in a lab activity, a badminton unit. This provided the students with practical application of many of the key concepts.

Data Collection

The researcher trained each participant for 1 hr about how to draw a concept map using a free online software program (i.e., http://cmap.ihmc.us/). The training was held in the computer lab so all participants had a chance to work on developing a concept map using this software. The software program helped participants easily draw and modify their concept map as well as export it as a JPEG file (i.e., image file) to send to the researcher via e-mail.

In this training, the researcher provided detailed information about a technical part of the software and shared examples of concept maps on different topics. For example, the researcher showed how to draw circles for concepts, how to make connections among circles using arrows, and how to put linking words on the arrows. The researcher also developed two concept maps using examples (i.e., tree, dogs) with the participants for them to have a sense of what the concept map looks like. The participants shared meaningful concepts about the examples, and the researcher drew the concept maps using the software so all participants could see the concept maps on the screen and be involved in this practice together.

The researcher encouraged the participants to think about concepts by answering focus questions. One of the focus questions was, what are dogs? The researcher had the participants answer this question using the concept map as if they were trying to explain a dog to the person who had never seen one. A different category of examples (i.e., concept links, cross-links, hierarchical links, and examples) was presented by the researcher. The participants had opportunities to ask questions about how to develop concept maps if they still did not understand clearly about what to do.

The researcher provided 30 key concepts and a focus question about assessment in PE before participants drew their own concept map. These 30 key concepts were derived from one of the assessment textbooks, titled as *Assessment-Driven Instruction in Secondary Physical Education: A Standards-Based Approach to Promoting and Documenting Learning* (Lund & Veal, 2013). This textbook was chosen because the authors have more than 32 years of teaching experience in PE including public schools and conducted the research and presented their works about assessment in PE (Lund & Veal, 2013). This textbook has a concept mapping exercise as a pre- and posttest as well. The focus question was, "how is assessment used by physical education teachers?" (Lund & Veal, 2013, p. ix). The lead researcher personally talked to one of the authors of this textbook and found out that they had added, deleted, and modified these key concepts over time and used the focus question and key concepts in developing the textbook.

Based on the given key concepts (i.e., 1–30) and the focus question, participants independently drew their own concept map using the online software. Participants in Group A (n = 30) drew one concept map during the first week of the class as a pretest and drew another concept map during the last week (i.e., Week 15) of the class in the same semester as a posttest. Between pre- and posttest, Group A received instructions about assessment in PE for one semester by the researcher. In the middle of the same semester, participants in Group B (n = 26) drew their concept map once without any instruction about assessment in PE. An example of a student concept map using all 30 key concepts is illustrated in Figure 1.

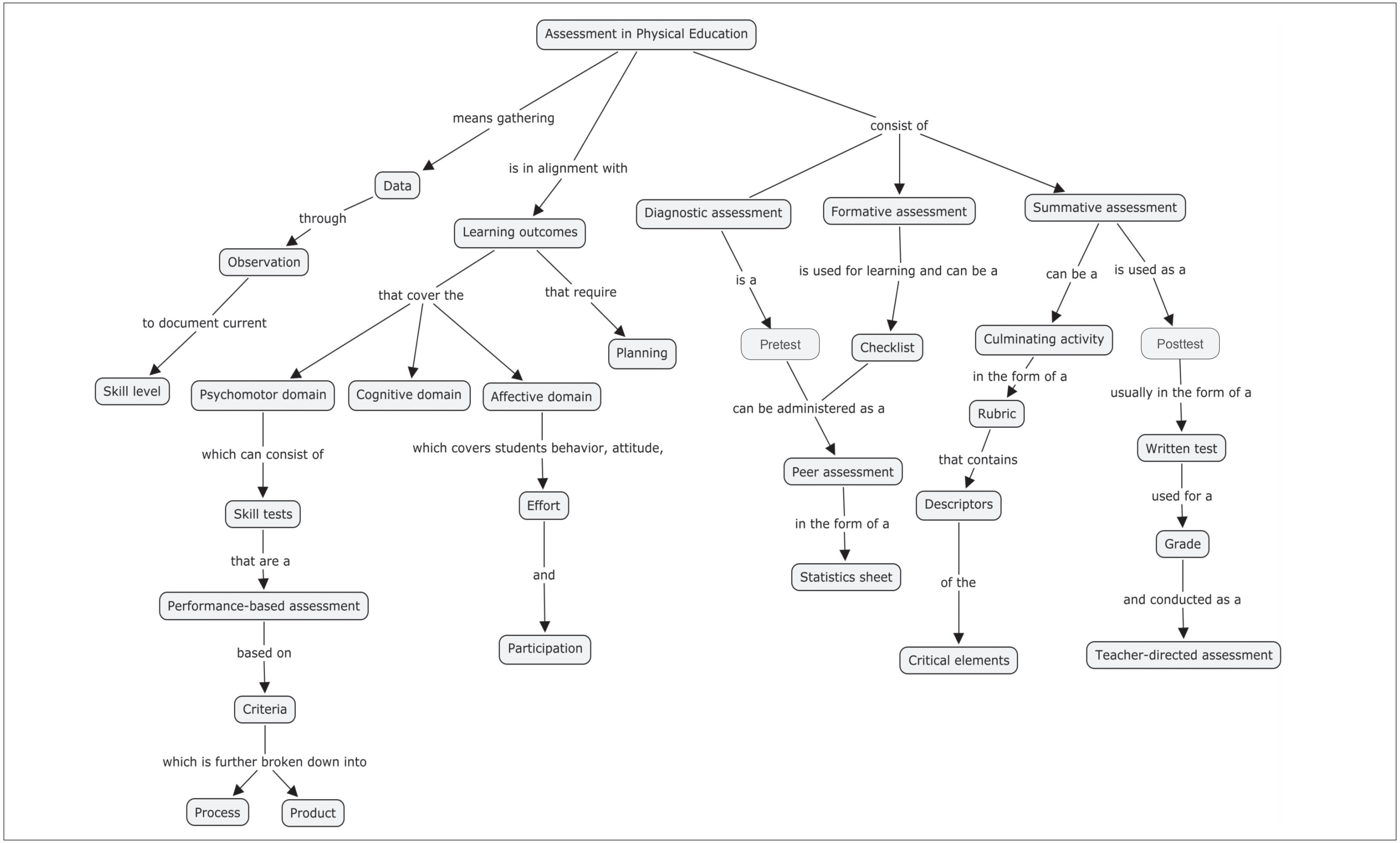


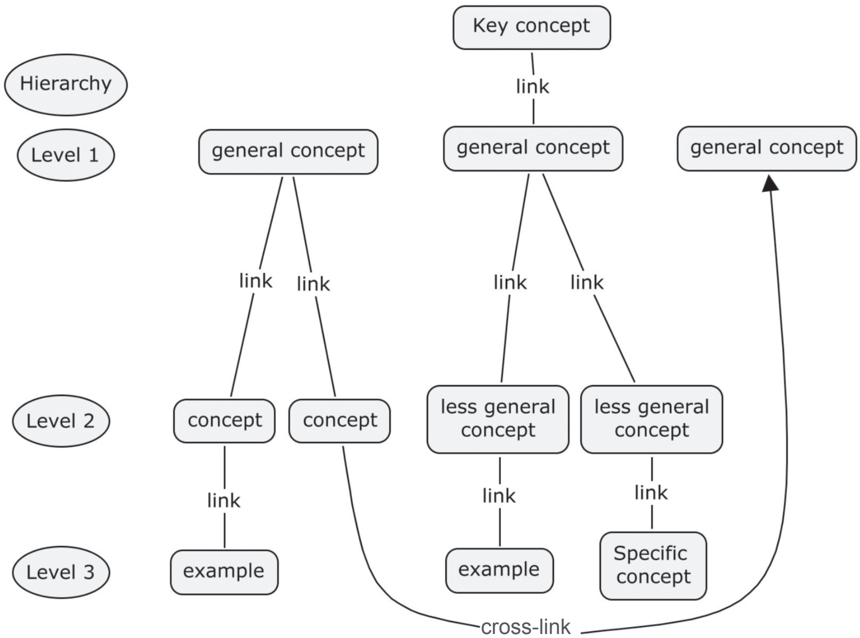
Figure 1. An example of student concept map using all 30 key concepts.

Scoring Methods

Two trained raters scored concept maps using a structural scoring method (Novak & Gowin, 1984) and a relational scoring method (McClure & Bell, 1990; McClure et al., 1999). Both raters are currently working in a PETE program at different universities and are knowledgeable about assessment in PE. As part of their training, they discussed the topic to make sure that they had a consensus in understanding the 30 key concepts. They also practiced together scoring using several examples of concept maps to be familiar with the scoring methods (i.e., structural and relational). They shared possible and valid concept links, cross-links, and hierarchical links. After the training, each rater scored 30% of the data independently using a structural scoring method (Novak & Gowin, 1984) and a relational scoring method (McClure & Bell, 1990; McClure et al., 1999). This 30% of the data was used to analyze interrater reliability for two scoring methods in concept maps.

Structural method. The structural scoring method is widely used in concept map assessment (Novak & Gowin, 1984) and is shown in Figure 2. The structural scoring method is used to score concept maps in four categories: concept links (1 point each), hierarchy (5 points each), cross-links (10 points each), and examples (1 point each). Any invalid links were given 0 points. Concept links indicate that participants link two concepts together using a line with a statement of the relationship. If a concept link was valid, one point was given. Hierarchy indicates that participants arrange one general concept at the top and one specific concept below and make a statement of the relationship. If a hierarchy was valid, it was given 5 points. Cross-links indicate that participants make connections between the concepts from different hierarchies and make a statement of the relationship. One cross-link was given 10 points if it was valid and significant. Examples indicate that participants can provide specific examples about the concepts. Each example was given 1 point if valid. Based on the structural scoring method, a total score and subscores for each category (i.e., concept links, cross-links, hierarchy, and examples) were recorded for each concept map.

Relational method. The relational scoring method was initially introduced by McClure and Bell (1990) and modified by McClure et al. (1999) from science education. According to McClure et al., the relational scoring method is used to score a concept map on a 4-point scale (i.e., 0 to 3 points). If the relationship between concepts was invalid, no point was given. If the relationship between



Scoring for this model:

Relationships (if valid)	1 × 8	= 8
Hierarchy (if valid)	2 × 5	= 10
Cross-links (if valid and significant)	1 × 10	= 10
Examples (if valid)	2 × 1	= 2
Total score		= 30

Figure 2. Model of structural scoring method.

concepts was valid but the statement (i.e., linking word or phrase) was incorrect, only 1 point was given. If the relationship between concepts was valid and the statement was correct, 2 points were given. If the relationship between concepts was valid and the direction of the arrow indicated a hierarchical, causal, or sequential relationship with a compatible statement, the highest points (i.e., 3 points) were given. Based on the relational scoring method, a total score and subscores for each point (i.e., 0 to 3 points) were recorded for each concept map.

Data Analysis

Statistical analyses were computed using the SPSS version 19.0 software package (SPSS, Inc., Chicago, IL, USA). To examine interrater reliability, intraclass correlation coefficient (ICC) was calculated using 30% of data. The known-group difference validity evidence of the concept map assessment was evaluated by comparing the total scores and subscores of the concept maps between two groups (i.e., Group A vs. Group B) using independent *t* tests. The responsiveness of the concept map assessment was computed by comparing the total scores and subscores of pre- and posttests from Group A using paired *t* tests. Cohen's *d* was calculated to determine the effect size. The diagram of the data analysis is shown in Figure 3.

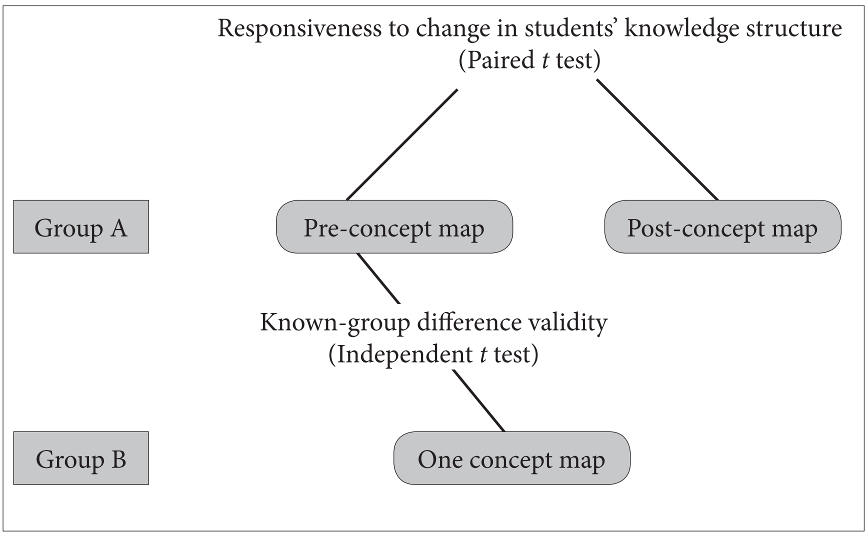


Figure 3. Diagram of data analysis.

Results

The validity and responsiveness of concept map assessment for both scoring methods were supported by the evidence of differences between Group A and Group B and between pre- and posttest for Group A in this study. Mean comparison of total scores and subscores between Group A and Group B and before and after instruction of Group A for structural scoring method is shown in Tables 1 and 2 and mean comparison for relational scoring method is shown in Tables 3 and 4.

Table 1
Mean Comparison of Total Scores and Subscores Between Group A and Group B for Structural Scoring Method

Component	Group A *M*(*SD*)	Group B *M*(*SD*)	*p*	Effect size (Cohen's *d*)
Concept Links	5.87 (4.15)	3.23 (2.86)	.009	0.74
Cross-Links	0 (0)	0 (0)	-	-
Hierarchy	0.17 (0.91)	0 (0)	.357	0.26
Examples	0 (0)	0 (0)	-	-
Total	6.03 (4.30)	3.23 (2.86)	.007	0.77

Table 2
Mean Comparison of Total Scores and Subscores Before and After Instruction of Group A for Structural Scoring Method

Component	Pretest *M*(*SD*)	Posttest *M*(*SD*)	*p*	Effect size (Cohen's *d*)
Concept Links	5.87 (4.15)	14.87 (5.49)	.001	1.85
Cross-Links	0 (0)	0.6 (2.30)	.163	0.37
Hierarchy	0.17 (0.91)	2.83 (3.87)	.001	0.95
Examples	0 (0)	0 (0)	-	-
Total	6.03 (4.30)	18.30 (8.01)	.001	1.91

Table 3
Mean Comparison of Total Scores and Subscores Between Group A and Group B for Relational Scoring Method

Scores	Group A *M*(*SD*)	Group B *M*(*SD*)	*p*	Effect size (Cohen's *d*)
1 Point	7.1 (5.09)	4.77 (2.72)	.035	0.57
2 Points	12.7 (9.3)	5.31 (4.96)	.001	0.99
3 Points	0.90 (2.25)	0.92 (2.21)	.969	0.00
Total	20.77 (8.53)	11 (6.78)	.001	1.27

Table 4
Mean Comparison of Total Scores and Subscores Before and After Instruction of Group A for Relational Scoring Method

Scores	Pretest *M*(*SD*)	Posttest *M*(*SD*)	*p*	Effect size (Cohen's d)
1 Point	7.1 (5.09)	7.17 (5.34)	.959	0.01
2 Points	12.7 (9.3)	27.97 (14.04)	.001	1.28
3 Points	0.90 (2.25)	3 (5.17)	.034	0.53
Total	20.77 (8.53)	38.13 (12.62)	.001	1.61

Structural Method

Interrater reliability for scoring the concept maps between two raters using structural method was very high (ICC = .985). Total scores for students in Group A (M = 6.03, SD = 4.30) were significantly higher than scores for students in Group B (M = 3.23, SD = 2.86), $t(54) = -2.83$, $p = .007$, which supports the known-group difference validity evidence. Cohen's d was .77, which indicates a medium to large effect size. For subscores for each category, concept links ($p = .009$) scores of Group A (i.e., senior group) were significantly different from scores for Group B (i.e., sophomore group).

Total scores of concept map assessment in Group A significantly increased after instruction from the mean of 6.03 (SD = 4.30) to the mean of 18.30 (SD = 8.01), $t(29) = -8.05$, $p < .001$, which supports the responsiveness of concept map assessment scores. Cohen's d was 1.91, which indicates a large effect size. The concept links (p = .001) and hierarchy (p = .001) subscores of concept maps increased significantly after instruction.

Relational Method

Interrater reliability for scoring the concept maps between two raters using relational method was very high (ICC = .946). Total scores for students in Group A (M = 20.77, SD = 8.53) were significantly higher than for students in Group B (M = 11.00, SD = 6.78), $t(54) = -4.69$, $p < .001$, which supports the known-group difference validity evidence. Cohen's d was 1.27, which indicates a large effect size. For subscores for each point, 1 point (p = .035) and 2 points (p = .001) subscores of Group A were significantly different from subscores of Group B.

Total scores of concept map assessment in Group A significantly increased after instructions from the mean of 20.77 ($SD = 8.53$) to the mean of 38.13 ($SD = 12.62$), $t(29) = -.8.55$, $p < .001$, which supports the responsiveness of concept map assessment scores. Cohen's d was 1.61, which indicates a large effect size. Two points ($p = .001$) and 3 points ($p = .034$) subscores of concept map assessment significantly increased after instruction.

Discussion

The results of this study provide evidence of validity and responsiveness of concept map assessment. It reflects predictable difference between senior students and sophomore students and the change in the students' knowledge structure before and after instruction in the PETE program. Because senior students took more major courses in the PETE program than sophomore students did, it was assumed that senior students had more knowledge about PE than sophomore students did. It is possible that senior students have more knowledge about assessment in PE than sophomore students do. It was also assumed that students' understanding or knowledge structure would change after instruction and thus higher concept map scores were expected after instruction than before instruction.

In this study, two scoring methods were used and both scoring methods supported the validity evidence for concept map assessment. These findings are different from West et al.'s (2002) findings, who reported that the relational scoring method did not show the resident doctors' knowledge changes before and after instructions and the knowledge differences based on their level of training. However, they demonstrated that the structural scoring method was a valid measure for concept map assessment. One possible explanation is that validity is context specific. Different educational settings may affect this difference (i.e., college students in PETE versus pediatric resident doctors in medical school). It is also possible that one scoring method may not be appropriate for a particular context. The small sample size ($n = 21$) of West et al.'s (2002) study could be another factor that influenced the results of their study. Therefore, the accumulation of validity evidence for concept map assessment is needed to further support the use of concept map assessment in PETE.

Based on the findings about the subscores of each category in structural method, there were no significant differences between Group A (i.e., senior student group) and Group B (i.e., sophomore student group) in the category of cross-links and hierarchy. One possible reason is because a limited number of key concepts (n = 30) were provided when the students developed their concept maps. With a limited number of key concepts, it may not be possible to make many cases for cross-links or hierarchy relationships. This result is the same as the findings of the relational scoring method in which no significant differences were observed in the scoring category of 3-point relationship between Group A and Group B.

Concept map assessment can be administered with or without a list of key concepts. Thirty key concepts about assessment in PE were provided in this study before the teacher candidates drew their concept maps. McClure et al. (1999) provided a list of key concepts from an educational psychology program when they asked students to produce a concept map. The selected key concepts, however, may influence their knowledge structure either positively or negatively depending on their level of knowledge about a particular topic. In some studies, in different subject matters, concepts were not provided to students (West et al., 2002; West et al., 2000). In this case, students needed to recall all important concepts about a given topic and to integrate their knowledge using the concepts about which they already knew when they drew their concept maps.

Limitations

The current study is not without limitations. Even though we provided a list of 30 key concepts selected by the experts, there may not have been enough key concepts about examples (i.e., one of scoring categories for structural scoring method), which limited the participants from making a case in their concept maps. We rarely found the category of examples and hierarchies from the concept maps in this study, which implies that the structural scoring method may not be appropriate if the list of key concepts were given. This could be the same for the relational scoring method. Due to a limited number of key concepts, not enough hierarchical, causal, or sequential relationships (i.e., 3-point relationships) were found in the relational scoring method. Because the scoring of the concept maps is intuitive, a high ICC was expected in both scoring methods (i.e., structural scoring method, ICC = .985; relational scoring method, ICC = .946). It may be better to use the total scores when the list of key concepts is provided. When the teacher wants to look at subscores of each category or each point, it may be better not to provide key concepts or to provide many key concepts.

Another limitation of this study is that we only established the known-group difference validity evidence and the responsiveness of the concept map assessment scores in PETE programs. The time and effort that teacher candidates spent to develop concept maps may have been an issue. Because concept mapping activities were conducted by students themselves in different conditions, some of them may not have done their best to develop their concept maps. Students also may have developed their post-concept map while looking at their pre-concept map. In this case, students' conceptual understanding could be compounded, even though the results of this study showed the changes of their knowledge structure.

Implications and Future Directions

Concept map assessment can be used to provide information about what and how to improve the contents of current courses within PETE programs. In the current study, there is evidence of validity and responsiveness of the concept map assessment scores in PETE programs. Concept map assessment can be an effective tool to evaluate PETE programs, as it may be used to measure teacher candidates' knowledge structure. It may be used as an ongoing self-assessment tool to measure what teacher candidates know and can do. It also allows teacher educators to identify teacher candidates' understandings and misunderstandings about a particular topic so they can modify or change their instruction. Furthermore, it helps teacher candidates become well-equipped and effective PE teachers.

PCK can be assessed using concept map assessment. Ayvazo and Ward (2011) demonstrated that PCK can be observed and measured using observation of student–teacher interactions and their appropriateness, which they called functional analysis. Even though direct observation using functional analysis of instructional adaptations may be used to measure teachers' PCK accurately, it does take much time to assess and analyze teachers' PCK, and it may not be an efficient way to measure their PCK. If more validation studies about concept map assessment are conducted in different contexts with different content, concept map assessment may be used as an alternative assessment tool to measure teacher candidates' PCK using an operational definition of PCK in PETE programs. In addition, it will help teacher candidates learn more PCK in their PETE programs.

Through concept map assessment, teacher candidates will obtain in-depth knowledge in PETE programs. Different objective tests are used in PETE programs including multiple-choice questions, true and false questions, short essay questions, journals, projects, and portfolios (Lund & Kirk, 2002). If alternative assessments including concept map assessment are used with these traditional tests, it would be better for teacher educators to help teacher candidates' learning. It is critical that different assessments be used for teacher candidates' learning, as well as their grade, in PETE programs.

Based on the findings and limitations of the current study, the following future directions for research are offered. First, future research is needed to compare conditions based on the amount of key concepts given (e.g., 30, 50, 80). The scores in subcategories of scoring methods need to be compared under conditions with different amounts of key concepts. In addition, future research would be to compare two conditions (i.e., providing key concepts versus not providing key concepts) and to examine how these conditions influence the effectiveness of concept map assessment in a particular content.

Second, future research is needed to establish different validity evidence for concept map assessment. For example, it may be possible to establish convergent validity in concept map assessment by comparing concept map assessment scores with standardized test scores in a particular content. Additional directions of future studies would be replication studies using different content, context, conditions, and participants. For example, concept map assessment could be used in other major courses in a PETE program such as motor behavior, curriculum, and adapted PE courses. It could also be used in the same content with different contexts (e.g., lecture-based course vs. lecture and field experience course).

Finally, future research could be conducted in a more controlled context to remove compounding factors. How much time participants spend and how much effort they put into developing concept maps may threaten the validity of concept map assessment scores. To remove these compounding factors, teacher educators could ask teacher candidates to develop their concept maps in a computer lab for a certain time so they spend the same amount of time developing their concept maps. This way, all teacher candidates experience the same environmental testing conditions.

References

Austin, L. B., & Shore, B. M. (1995). Using concept mapping for assessment in physics. *Physics Education, 30*, 41–45.

Ayvazo, S., & Ward, P. (2011). Pedagogical content knowledge of experienced teachers in physical education: Functional analysis of adaptations. *Research Quarterly for Exercise and Sport, 82*, 675–684.

Besterfield-Sacre, M., Gerchak, J., Lyons, M., Shuman, L., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education, 93*, 105–115.

Ennis, C. D., Mueller, L. K., & Zhu, W. (1991). Description of knowledge structures within a concept-based curriculum framework. *Research Quarterly for Exercise and Sport, 62*, 309–318.

Kassab, S. E., & Hussain, S. (2010). Concept mapping assessment in a problem-based medical curriculum. *Medical Teacher, 32*, 926–931.

Lund, J. L., & Kirk, M. F. (2002). *Performance-based assessment for middle and high school physical education*. Champaign, IL: Human Kinetics.

Lund, J. L., & Veal, M. L. (2013). *Assessment-driven instruction in secondary physical education: A standards-based approach to promoting and documenting learning*. Champaign, IL: Human Kinetics.

McClure, J. R., & Bell, P. E. (1990). *Effects of an environmental education related STS approach instruction on cognitive structures of preservice science teachers*. Retrieved from ERIC. (ED341582)

McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching, 36*, 475–492.

Mohammed, A. M. S. E. (2010). The effect of using concept maps strategy on the maintenance level in kinetic education for grade one female students in faculty of physical education for girls in Elgezira. *World Journal of Sport Sciences, 3*, 119–124.

National Association for Sport and Physical Education. (2008). *National initial physical education teacher education standards*. Retrieved from http://www.aahperd.org/naspe/grants/accreditation/upload/2008-National-Initial-PETE-Standards.pdf

National Council for Accreditation of Teacher Education. (2008). Professional standards for the accreditation of schools, colleges, and departments of education. Retrieved from http://www.ncate.org/Standards/NCATEUnitStandards/UnitStandardsinEffect2008/tabid/476/Default.aspx

Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them* (Technical Report IHMC Cmaps Tools 2008-1). Pensacola, FL: Florida Institute for Human and Machine Cognition.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York, NY: Cambridge University Press.

Pearsall, N. R., Skipper, E. J., & Mintzes, J. J. (1997). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education, 81*, 193–215.

Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching, 35*, 1103–1127.

Rink, J. E., French, K., Lee, A. M., Solomon, M. A., & Lynn, S. K. (1994). A comparison of pedagogical knowledge structures of preservice students and teacher educators in two institutions. *Journal of Teaching in Physical Education, 13*, 140–162.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33*, 569–600.

Rye, J. A., & Rubba, P. A. (2002). Scoring concept maps: An expert map-based scheme weighted for relationships. *School Science and Mathematics, 102*, 33–44

West, D. C., Park, J. K., Pomeroy, J. R., & Sandoval, J. H. (2002). Concept mapping assessment in medical education: A comparison of two scoring systems. *Medical Education, 36*, 820–826.

West, D. C., Pomeroy, J. R., Park, J. K., Gerstenberger, E. A., & Sandoval, J. H. (2000). Critical thinking in graduate medical education: A role for concept mapping assessment? *Journal of the American Medical Association, 284*, 1105–1110.