The Physical Educator Vol. 79 • pp. 569–585 • 2022
https://doi.org/10.18666/TPE-2022-V79-I5-10958

SPORT

# Validating the Volleyball Common Content Knowledge Test

*Emi Tsuda, Phillip Ward, Debra Sazama, Yaohui He, Harry Lehwald, Bomna Ko, José A. Santiago, Xiuye Xie*

## Abstract

*The purpose of this study was to create a valid and reliable volleyball common content knowledge (VB-CCK) test in secondary physical education contexts in the United States. Two physical education teacher educators served as content experts and developed test items for the VB-CCK test. We then established content validity with a group of in-service teachers and determined face validity with preservice teachers. The Rasch model was used in data analysis. The standardized residual contrasts for the items loaded in the range of -.4 to .4, confirming the test was unidimensional and measured VB-CCK. Both infit mean-square residuals and outfit mean-square residuals were within the criterion range of .5 to 1.5 (infit = .8 to 1.1; outfit = .8 to 1.5), which showed the range of the question difficulties matched the respondents' knowledge levels. The item separation index was 3.46 (reliability = .92; an excellent level) and the person separation index was 1.55 (reliability = .71; an acceptable level). The results provided evidence of the range of question difficulties and respondents' knowledge levels. The Wright Map added additional evidence that the 40 questions were well distributed from easy to difficult levels and could discriminate different knowledge levels of preservice teachers' VB-CCK. The study was successful in developing*

Emi Tsuda, School of Sport Sciences, West Virginia University. Phillip Ward, Department of Human Sciences, The Ohio State University. Debra Sazama, Exercise and Sport Science Department, University of Wisconsin–La Crosse. Yaohui He, College of Physical Education and Health, East China Normal University. Harry Lehwald, Department of Human Sciences, The Ohio State University. Bomna Ko, Department of Kinesiology, East Carolina University. José A. Santiago, Department of Kinesiology, Sam Houston State University. Xiuye Xie, Department of Kinesiology, Missouri State University. Please send author correspondence to emi.tsuda@mail.wvu.edu

*a valid and reliable instrument for PETE programs to assess secondary VB-CCK of their preservice teachers. We propose future CCK studies for both elementary content and secondary content.*

Common content knowledge (CCK) is one of the significant knowledge areas impacting teaching effectiveness (Kim et al., 2018; Ward & Ayvazo, 2016). CCK includes the knowledge needed to perform a sport or activity and knowledge of rules, safety, etiquette, techniques, and tactics (Ward, 2009). The National Standards for Initial Physical Education Teacher Education recognize the importance of CCK, with Standard 1.a indicating that "*candidates will describe and apply CCK for teaching preK–12 physical education*" (SHAPE America, 2017, p.1). As such, physical education teacher education (PETE) programs are required to demonstrate preservice teachers' (PSTs) CCK levels in various content areas to describe their programs' quality to educate effective teachers.

However, in the history of the PE field, there have been few valid and reliable knowledge tests for U.S. PETE programs to adopt. One book that PETE programs have used is *Test Questions for Physical Education Activities* (McGee & Farrow, 1987). The book is now both out of date and out of print. It provided tests for a variety of sports and physical activities. Most PETE programs use faculty-created content-valid tests to assess PSTs' knowledge in lieu of valid and reliable tests. There is a pressing need for valid and reliable CCK tests because they demonstrate that PETE is an evidence-based process and that PETE programs teach the required CCK for PSTs who will work in preK–12 schools.

Research on CCK has focused primarily on health-related fitness knowledge (Barnett & Merriman, 1994; Castelli & Williams, 2007; Ince & Hünük, 2013; Miller & Housner, 1998; Santiago et al., 2012; Santiago & Morrow, 2021). Until recently, few studies have examined sports-related content, which represents a significant portion of the content areas in the K–12 physical education standards (SHAPE America, 2014). Studies have demonstrated the reliability of CCK tests for PSTs for soccer in China (He et al., 2018), South Korea (Lee et al., 2018), and Turkey (Dervent et al., 2018). Tests have also been developed for volleyball in Japan (Tsuda, Ward, Yoshino, et al., 2019b) and gymnastics in Turkey (Devrilmez et al., 2019).

A similar process was used in the development of the extant CCK tests. First, content experts with extensive performing and teaching/coaching experiences of a sport, who also understood secondary school physical education and PETE programs' contexts, developed test questions. Next, in-service teachers content-validated the test questions and answers. Then, the tests were pilot tested with a group of PSTs to evaluate the face validity. Finally, the tests were administered to PSTs, with sample sizes ranging from 126 (Tsuda, Ward, Yoshino, et al., 2019) to 530 (He et al., 2018). Rasch measurement modeling (Rasch, 1980) was typically utilized in the analysis of validity and reliability of the tests.

Rasch measurement is a valuable alternative to classical test theory. It is considered the gold standard for test development because of its ability to overcome shortcomings in current measurement, particularly negative skewness and limited discrimination (Edelsbrunner & Dablander, 2019). Rasch techniques allow researchers to make critical corrections when using raw test score data. A notable strength of Rasch techniques is they allow nonlinear raw data to be converted to a linear scale, which can then be evaluated through the use of parametric statistical tests (Boone, 2016). Rasch analyses can also be used in the investigation of other important instrumentation issues (e.g., step ordering/step disordering, item reliability, person reliability, differential item functioning, and differential test functioning; Boone, 2016).

Though there is a developing bank of CCK tests, tests do not always translate across countries. Each country has a different educational context that creates different expectations for teachers. We found one CCK test for tennis in the United States (Tsuda et al., 2021). Developing CCK tests for PETE in the United States is important because the scope of content knowledge is contextually specific. He et al. (2018) reported that in the development of their soccer tests, Chinese scholars initially used CCK test questions developed in the United States. However, the test questions did not adequately represent the desired population of possible questions for Chinese physical educators, and they needed to develop test questions that covered the content taught in secondary physical education in China, which was more in-depth in the content than the questions that they received from the United States.

Developing context-specific, valid, and reliable CCK tests is vital in the establishment of teacher knowledge and the assessment of the impact of the instruction. The purpose of this study was to create a valid and reliable volleyball CCK (VB-CCK) test in secondary physical education contexts in the United States. The goals of this study were to (a) establish the content and face validity of the test of VB-CCK and (b) assess its reliability and validity. Rasch measurement modeling was utilized in this study.

## Method

Institutional review board approval and informed consent from participants was obtained for the study. Volleyball was selected as the content area on the basis of informal interviews with 10 faculty members in PETE programs because volleyball is one of the commonly taught content areas in secondary physical education. Teaching volleyball in secondary physical education allows teachers to address the grade-level outcomes for net/wall games and lifetime activities in both middle school physical education and high school physical education (SHAPE America, 2014; Ward & Lehwald, 2018).

### Developing the Volleyball Common Content Knowledge Test for Preservice Teachers

Two physical education teacher educators served as content experts and developed test questions for the VB-CCK test. One expert coached volleyball at the collegiate level and in youth community programs and taught physical education at the high school level. The other expert taught volleyball in middle and high school physical education programs and published scholarly works in volleyball. They developed the VB-CCK test by determining the scope of the test and then creating the test items. We then established content validity and determined face validity.

#### *Determining the Scope of the Test*

The content experts developed a table of specifications to determine the difficulty and distribution of the questions. The weighted difficulty of the questions was designed to be 30% easy (people who have limited knowledge in volleyball can answer), 40% moderate, and 30% hard (only volleyball experts can answer). The item distribution was determined to ensure different difficulty of items were equally included in the test to discriminate various knowledge levels of PSTs. The experts used a volleyball content map (i.e., graphic organizers of content to be taught in schools across grade levels) presented in *Effective Physical Education Content and Instruction* (Ward & Lehwald, 2018) as a guide for this process. The content map presented by Ward and Lehwald (2018) has been validated and aligns with the National Standards and Grade Outcomes for K–12 Physical Education (SHAPE America, 2014). In addition, we referenced *Teaching Volleyball: Steps to Success* (Viera & Ferguson, 1989) to ensure the accuracy of the content knowledge. The table of specifications was modified until the two experts reached a consensus.

### *Developing Test Items*

Using the table of specifications, the content experts developed test items. They revised questions until reaching agreement. Forty multiple-choice questions were developed. While having more test items could have improved the test validity and reliability, we decided on this total item number, with practical consideration (i.e., the fatigue of the test takers). Each item included a stem and four responses, one correct answer, and three distractors. Items for rules and safety covered knowledge such as court orientation, the scoring system, the number of contacts per side, player positions, and centerline violation. Questions for techniques covered the critical elements of forearm passing, setting, underhand and overhand serving, attacking, and blocking. Questions for tactics covered knowledge such as when to set or pass, when to block, and when to serve. The experts decided to exclude advanced volleyball knowledge (e.g., jump service, offensive plays, and defensive schemes) because this volleyball test aimed to provide an essential knowledge base to teach beginning volleyball learners in secondary school physical education settings.

### *Establishing Content Validity*

To establish content validity, we identified six in-service teachers who were recognized as effective teachers by their colleagues and PETE faculty and who had instructed volleyball in secondary physical education. The six teachers reviewed the test and provided feedback on (a) the questions and answers relative to what they taught to secondary students (Grades 6 to 12, ages 12 to 18) and (b) whether the items and answers were developmentally appropriate for secondary students. The teachers provided feedback on the clarity of test items, and the content experts revised the items based on the feedback.

### *Weighting Assessment and Securing Face Validity*

Finally, 11 PSTs who we preestablished as having varying volleyball knowledge (i.e., low, medium, and high) were recruited and took the VB-CCK test for pilot testing. We determined the knowledge levels through years of volleyball-playing experiences as the proxy measure. The PSTs who varied in knowledge levels were invited to determine if the PSTs with low knowledge would score low, and those with average knowledge level would have medium scores, and those with most knowledge would have high scores on the test. For the establishment of face validity, the PSTs also examined the clarity of test items in this process. The pilot-testing results provided evidence that the test could distinguish different knowledge levels ($M = 52.72$; range 39.53 to 85.00). The content experts also improved the clarity of the questions on the basis of the feedback.

## Test Administration Process and Participants

Once the content validity and the face validity of the VB-CCK test were established, we administered it to a group of physical education PSTs to examine the validity and reliability of the measure. Participants were 73 ($n_{male} = 50$; $n_{female} = 23$) physical education PSTs from three universities. The participants had various levels of experience in playing volleyball. Before taking the VB-CCK test, the PSTs completed a demographic questionnaire to provide information on their volleyball backgrounds, including years of playing, teaching/coaching experiences, and courses completed in their PETE programs. The participants also indicated their perceived volleyball knowledge levels and confidence in playing and teaching/coaching volleyball, using the 10-point Likert scale (1 = *very little knowledge/confidence*; 10 = *very high knowledge/confidence*). Table 1 provides a summary of the participants' volleyball backgrounds. The test was implemented with a hard copy in an in-person setting. The results of the test were entered into an Excel spreadsheet and analyzed.

## Data Analysis

The Rasch model (Linacre, 2011; Rasch, 1980) was used for data analysis. The Rasch model maximizes the homogeneity of the test items and reduces the redundancy of the items by decreasing items and/or scoring levels to produce a valid and simple measure (Granger, 2008). The Rasch model is underpinned by the assumption that the probability of a person's response to a question depends on the difficulty of the item and the person's ability (Linacre, 2011). Measures constructed through the Rasch model are unidimensional and have a range of difficulty within the measure (Linacre, 2011). In this study, we analyzed the validity and reliability of the volleyball test through four dimensions of the Rasch model: (a) unidirectionality, (b) model-data fit, (c) item/person reliability and separation, and (d) a Wright Map (Linacre, 2011; Rasch, 1980).

### *Unidimensionality*

The Rasch model's fundamental assumption is that the test questions measure one construct rather than multiple constructs, namely, unidimensionality (Linacre, 2011). The construct in this study was the CCK of volleyball. The unidimensionality was judged through standardized residual contrast loading. If the results of the standardized residual contrast one plot are in the range of -.4 to .4, this would confirm the unidimensionality of the test.

### *Model–Data Fit*

The model–data fit analysis examines the alignment of the question difficulties and respondents' knowledge levels. We utilized the results of infit and outfit statistics to explore the model–data fit. Infit statistics ensure that only higher-knowledge respondents could answer difficult questions and participants who have limited knowledge in volleyball could answer easy questions. Outfit statistics look at the unexpected results for respondents. For example, the question about the critical elements of the double block, which is considered a difficult question, would be answered only by higher-knowledge participants (infit), not the lower-knowledge-level respondents (outfit).

To examine the infit and outfit of the measure, we utilized mean-square residuals (MNSQ) and standardized mean-square residuals (ZSTD). The MNSQ is a squared residual based on the difference between the observed response patterns and the predicted response patterns, which is a chi-squared calculation. The ZSTD is a normalized $t$ score of the residual and provides $t$-test statistics measuring the probability of the MNSQ calculation occurring by chance (Liu,

**Table 1**

*Demographic Background of the Participants ($n = 73$; $n_{male} = 50$, $n_{female} = 23$)*

| Years of volleyball playing experiences | | | Taking volleyball class scale | | 10 point-Likert scale | | | |
|---|---|---|---|---|---|---|---|---|
| **Years** | **Playing** | **Teaching/ coaching** | | | **Scale** | **Volleyball knowledge level** | **Confidence in playing volleyball** | **Confidence in teaching/ coaching volleyball** |
| None | 51 | 60 | Yes | 25 | 1–2 | 7 | 5 | 27 |
| <1–3 | 16 | 13 | No | 48 | 3–4 | 19 | 14 | 20 |
| 4–7 | 3 | 0 | | | 5–6 | 18 | 19 | 12 |
| 8+ | 3 | 0 | | | 7–8 | 23 | 19 | 9 |
| | | | | | 9–10 | 6 | 16 | 5 |

2010). Our focus was on an MNSQ acceptable range of .5 to 1.5 since the ZSTD values are calculated through the MNSQ (Linacre, 2011). ZSTD values with a range of -1.9 to 1.9 were examined only when MNSQ values were outside of the acceptable range (Linacre, 2011).

### *Item–Person Separation and Reliability*

The results of the separation index demonstrate the range of question difficulties (i.e., an item separation index) and respondents' knowledge levels (i.e., a person separation index). The index value of 1.5 is an acceptable level of separation, 2.0 is a good level of separation, and 3.0 shows an excellent level of separation (Bond & Fox, 2007). A reliability index shows a degree of confidence for either item or person separation, with 1.0 being the most confident, and as the value goes higher, it indicates the test could distinguish more knowledge levels (i.e., .9 or higher = 3 or more levels, .8 = 2 or 3 levels, .7 = 1.5 to 2 levels, .6 = 1 to1.5 levels; Linacre, 2011). The reliability value less than .5 implies that the difference between measures is primarily due to measurement error and not acceptable.

### *Wright Map*

A Wright Map is a visual representation of item difficulties and performance distribution of respondents (Linacre, 2011). The right side of the map illustrates the highest ability scores on the uppermost section and the lowest ability score at the lowest. On the left side of the map, the most difficult item is plotted on the top and the easiest item on the bottom. The ideal test has the items and respondents distributed from the highest to the lowest on the Wright Map.

## Results

The standardized residual contrasts for the items loaded in the range of -.4 to .4, and this confirmed the test was unidimensional and measured volleyball CCK. Table 2 shows the results of MNSQ and ZSTD for each test item, which shows the alignment of the item difficulties and the knowledge levels of the respondents. Both infit MNSQ and outfit MNSQ ranged within the criterion range of .5 to 1.5 (infit = .8 to 1.1; outfit = .8 to 1.5), which showed that the range of the question difficulties matches the knowledge levels of the respondents.

Table 3 illustrates the results of the item and person separation index and reliability. The item separation index was 3.46 (reliability = .92; an excellent level) and the person separation index was 1.55 (reliability = .71; an acceptable level). The results provided evidence of the range of question difficulties and knowledge levels of the respondents. The Wright Map added additional evidence that the 40 questions were well distributed from easy to difficult levels and could discriminate different knowledge levels of PSTs' VB-CCK (Figure 1).

**Table 2**
*Item Fit of Rasch Analysis for the Volleyball Common Content Knowledge Measure for Preservice Teachers*

| | | | Infit | | Outfit | | PT-measure |
|---|---|---|---|---|---|---|---|
| **Item** | **Difficulty** | ***SE*** | **MNSQ** | **ZSTD** | **MNSQ** | **ZSTD** | **corr.** |
| 1 | -.22 | .25 | 1.12 | 1.68 | 1.16 | 1.73 | .14 |
| 2 | -.47 | .25 | 1.13 | 1.76 | 1.12 | 1.31 | .14 |
| 3 | -.72 | .25 | 1.07 | .83 | 1.22 | 1.94 | .17 |
| 4 | -1.83 | .31 | .88 | -.61 | .81 | -.71 | .41 |
| 5 | 1.06 | .29 | 1.02 | .16 | 1.10 | .55 | .23 |
| 6 | 1.33 | .31 | 1.18 | 1.01 | 1.29 | 1.16 | .01 |
| 7 | 1.43 | .32 | 1.19 | .98 | 1.44 | 1.58 | -.04 |
| 8 | -1.56 | .29 | .83 | -1.14 | .83 | -.79 | .48 |
| 9 | -1.48 | .28 | 1.19 | 1.36 | 1.57 | 2.54 | -.09 |
| 10 | -2.38 | .36 | .96 | -.05 | 1.10 | .39 | .23 |
| 11 | -1.04 | .26 | 1.04 | .39 | 1.20 | 1.39 | .20 |
| 12 | .67 | .27 | 1.05 | .43 | 1.07 | .50 | .23 |
| 13 | .27 | .25 | .94 | -.68 | .94 | -.53 | .39 |
| 14 | .21 | .25 | .89 | -1.27 | .85 | -1.47 | .47 |
| 15 | .67 | .27 | .96 | -.27 | 1.01 | .14 | .33 |
| 16 | .82 | .27 | 1.03 | .27 | 1.01 | .10 | .26 |
| 17 | 1.89 | .37 | 1.15 | .64 | 1.56 | 1.48 | -.05 |
| 18 | .90 | .28 | .99 | -.04 | 1.08 | .49 | .28 |
| 19 | .53 | .26 | .92 | -.73 | .86 | -1.04 | .43 |
| 20 | -.04 | .25 | .87 | -1.89 | .86 | -1.57 | .50 |
| 21 | .60 | .26 | 1.19 | 1.62 | 1.27 | 1.76 | .03 |
| 22 | .27 | .25 | .87 | -1.48 | .83 | -1.65 | .50 |

**Table 2 (cont.)**

| Item | Difficulty | *SE* | Infit MNSQ | Infit ZSTD | Outfit MNSQ | Outfit ZSTD | PT-measure corr. |
|---|---|---|---|---|---|---|---|
| 23 | -.65 | .25 | 1.04 | .47 | 1.14 | 1.30 | .23 |
| 24 | 1.06 | .29 | 1.08 | .56 | 1.43 | 1.93 | .08 |
| 25 | -.22 | .25 | .97 | -.42 | .96 | -.43 | .36 |
| 26 | .53 | .26 | 1.04 | .44 | 1.08 | .64 | .24 |
| 27 | .47 | .26 | 1.00 | .02 | .99 | -.06 | .31 |
| 28 | -.35 | .25 | 1.01 | .17 | .98 | -.17 | .31 |
| 29 | -1.33 | .27 | .86 | -1.12 | .76 | -1.45 | .49 |
| 30 | -1.11 | .26 | .84 | -1.60 | .76 | -1.68 | .52 |
| 31 | 1.06 | .29 | .86 | -.90 | .75 | -1.28 | .49 |
| 32 | -2.38 | .36 | .88 | -.39 | .63 | -1.08 | .42 |
| 33 | .21 | .25 | .98 | -.15 | .97 | -.30 | .34 |
| 34 | .34 | .26 | .86 | -1.61 | .83 | -1.52 | .51 |
| 35 | .08 | .25 | .96 | -.52 | .98 | -.17 | .36 |
| 36 | .67 | .27 | 1.11 | .92 | 1.21 | 1.36 | .13 |
| 37 | -.72 | .25 | .98 | -.16 | .99 | -.09 | .32 |
| 38 | .47 | .26 | .98 | -.16 | 1.02 | .18 | .32 |
| 39 | .75 | .27 | .99 | -.06 | 1.10 | .67 | .28 |
| 40 | .21 | .25 | .82 | -2.26 | .77 | -2.38 | .57 |
| *M* | .00 | .27 | .99 | -.1 | 1.04 | .1 | |
| P.*SD* | 1.01 | .05 | .11 | 1.0 | .22 | 1.2 | |

*Note.* MNSQ = mean-square residual; ZSTD = standardized mean-square residual.

**Table 3**
*Summary Statistics of the Rasch Analysis for the Volleyball Common Content Knowledge Measure for Preservice Teachers*

| Statistic | Item | Person |
|---|---|---|
| Number measured | 40 | 73 |
| Location *M* (*SD*) | 32.5 (14.7) | 17.8 (5.4) |
| Infit MNSQ *M* (*SD*) | .99 (-.11) | 1.00 (.17) |
| Outfit MNSQ *M* (*SD*) | 1.04 (.22) | 1.04 (.29) |
| Reliability | .92 | .71 |
| Separation | 3.46 | 1.55 |

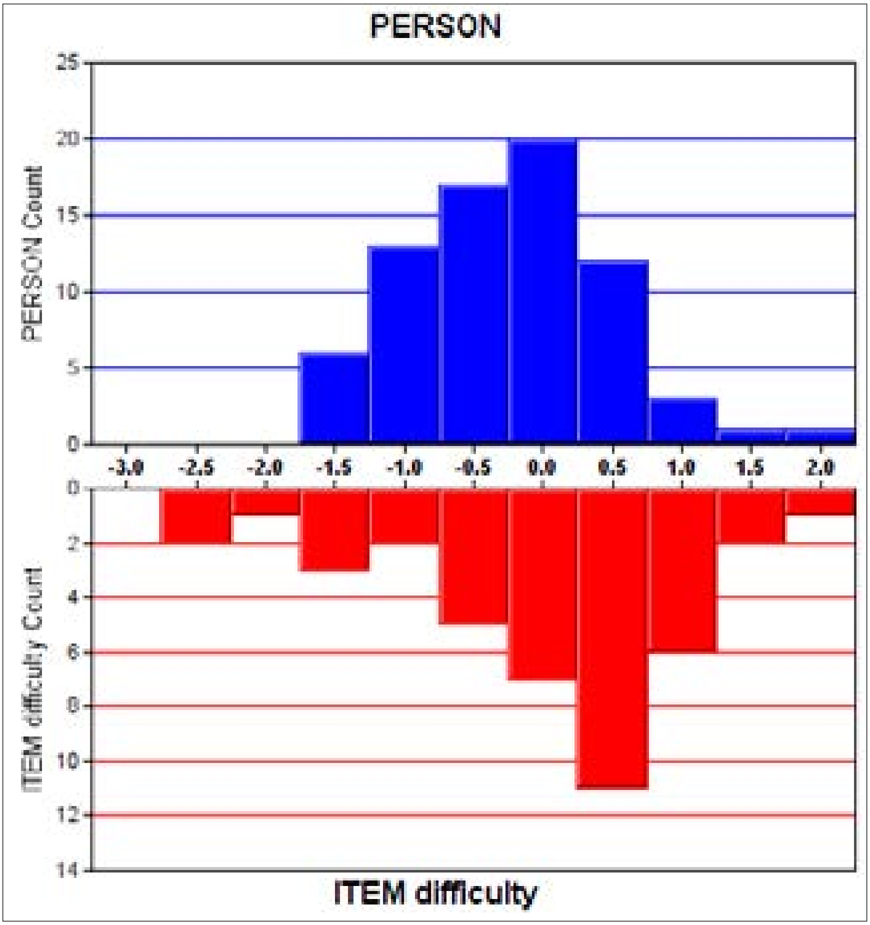
## Discussion

Few CCK tests measure PSTs' essential knowledge bases to teach secondary physical education in the United States, and this study was designed for the creation of one for volleyball. Content validity was established for secondary physical education. Rasch analysis confirmed that the VB-CCK test of PSTs measured a unidimensional construct. The Wright Map demonstrated the test had a good range of the question difficulties to discriminate a wide range of knowledge levels of respondents for secondary (Grades 6 to 12) physical education. In addition, Rasch analysis indicated the VB-CCK test was a valid and reliable measure of PSTs' volleyball CCK to teach secondary physical education.

The VB-CCK test can be used to help teacher educators define the knowledge to guide their instruction and ensure that their instruction is aligned with the expected learning outcomes and the National Standards and Grade-Level Outcomes for K–12 Physical Education (SHAPE America, 2014). Physical education teacher educators need robust assessments if the field moves further toward evidence-based practice and the provision of evidence that they teach what they claim. Making assessments such as the VB-CCK test available is one step in that direction.

As we noted in the introduction, there is a clear need for the development of more CCK tests in the United States. To that end, we have several recommendations. We suggest the continued creation of CCK tests in different content areas for both elementary schools and secondary schools for the creation of a bank of tests for teacher education. Physical education contains a wide range of content areas. SHAPE America (2014) grade-level outcomes identify numbers of skills to be taught in elementary physical education including locomotors, nonlocomotors or stability, and manipulative skills tied to in the areas of games skills, educational gymnastics, and dance (SHAPE America, 2014). For middle and high schools, SHAPE America (2014) grade-level outcomes categorize physical education content in seven areas: outdoor pursuits, fitness activities, dance and rhythmic activities, aquatics, individual-performance activities, games and sports, and lifetime activities. Considering the importance of CCK in effective instruction, it would be ideal to have a valid and reliable CCK test for each content area.

**Figure 1**
*Wright Map for the Volleyball Common Content Knowledge Measure for Preservice Teachers*




We also suggest the use of Rasch modeling for the test development because it is a gold standard for these sorts of assessments and has routinely demonstrated its utility (Dervent et al., 2018; Devrilmez et al., 2019; He et al., 2018; Lee et al., 2018; Tsuda, Ward, Li, et al., 2019) in and outside of the field (Panayides et al., 2010). As we noted earlier, Rasch techniques allow nonlinear raw data to be converted to a linear scale, which can then be evaluated through the use of parametric statistical tests, which have a greater statistical power (Boone, 2016). Rasch analysis also allows for the investigation of multiple testing issues (item reliability, person reliability, different item functioning, different test functioning; Boone, 2016). Further, we recommend the use of Rasch analysis beyond its current use in the assessment of CCK to other assessment areas in PETE that demonstrate PSTs' knowledge. Collective efforts of physical educators enable the development of resources that could improve the field.

## Limitations

There are three limitations in this study. First, the test was validated in the context of the United States with PSTs in PETE programs. As such, the test needs to be used cautiously in other contexts because the depth of CCK needed to teach secondary physical education differs in different educational contexts (He et al., 2018). A second limitation is the coverage of this test may not exactly align with what is taught in each PETE program in the United States. Each teacher educator could make a judgment on what questions to include or exclude when adopting this test in their instruction. The number of questions was 40. Having a larger number of questions could increase the coverage of the content and enhance the reliability of the test. However, with the considerations of fatigue of the test takers, we judged that having 40 questions was appropriate.

## Conclusion

The study was successful in developing a valid and reliable measure for PETE programs to assess secondary VB-CCK of their PSTs. From a methodological standpoint, this study demonstrates the utility of the Rasch methodology for test development for PETE. Our recommendation includes the use of it beyond its current use in the assessment of CCK. This study represents a step in the direction of broadening evidence-based practice of PST knowledge, specifically concerning particular interpretations about PSTs' CCK preparation. We propose future CCK studies for both elementary content and secondary content.

# References

Barnett, B. E., & Merriman, W. J. (1994). Knowledge of physical fitness in prospective physical education teachers. *Physical Educator*, *51*(2), 74–77.

Bond, T. G., & Fox, M. T. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Castelli, D., & Williams, L. (2007). Health-related fitness and physical education teachers' content knowledge. *Journal of Teaching in Physical Education*, *26*(1), 3–19. https://doi.org/10.1123/jtpe.26.1.3

Dervent, F., Devrilmez, E., Ince, M. L., & Ward, P. (2018). Reliability and validity of football common content knowledge test for physical education teachers. *Hacettepe Journal of Sport Sciences*, *29*(1), 39–52.

Devrilmez, E., Dervent, F., Ince, M. L., & Ward, P. (2019). A test of common content knowledge for gymnastics. *European Physical Education Review*, *25*(2), 512–523. https://doi.org/10.1177/1356336X17751232

Edelsbrunner, P. A., & Dablander, F. (2019). The psychometric modeling of scientific reasoning: A review and recommendations for future avenues. *Educational Psychology Review*, *31*, 1–34. https://doi.org/10.1007/s10648-018-9455-5

Granger, C. (2008). Rasch analysis is important to understand and use for measurement. *Rasch Measurement Transactions*, *21*(3), 1122–1123.

He, Y., Ward, P., & Wang, X. (2018). Validation of a common content knowledge test for soccer. *Journal of Teaching in Physical Education*, *37*(4), 407–412. https://doi.org/10.1123/jtpe.2017-0204

Ince, M. L., & Hünük, D. (2013). Experienced physical education teachers' health-related fitness knowledge level and knowledge internalization processes. *Education and Science*, *38*(168), 304–317.

Kim, I., Ward, P., Sinelnikov, O., Ko, B., Iserbyt, P., Li, W., & Curtner-Smith, M. (2018). The influence of content knowledge on pedagogical content knowledge: An evidence-based practice for physical education. *Journal of Teaching in Physical Education*, *37*(2), 133–143. https://doi.org/10.1123/jtpe.2017-0168

Lee, H. J., Ko, B., Ward, P., Lee, K., Higginson, K., He, Y., & Lee, Y. S. (2018). Development and validation of common content knowledge test of soccer. *Korean Journal of Sport Science*, *29*(4), 650–662. https://doi.org/10.24985/kjss.2018.29.4.650

Linacre, J. M. (2011). *A user's guide to Winsteps Ministep Rasch-Model computer program.* Winsteps.

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. Information Age Publishing.

McGee, R., & Farrow, A. (1987). *Test questions for physical education activities*. Human Kinetics.

Miller, M. G., & Housner, L. (1998). A survey of health-related physical fitness knowledge among pre-service and in-service physical educators. *Physical Educator*, *55*(4), 176–186.

Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*, *36*(4), 611–626. https://doi.org/10.1080/01411920903018182

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Santiago, J. A., Disch, J. G., & Morales, J. (2012). Elementary physical education teachers' content knowledge of physical activity and health-related fitness. *Physical Educator*, *69*(4), 395–412.

Santiago, J. A., & Morrow, J. R. (2021). A study of pre-service physical education teachers' content knowledge of health-related fitness. *Journal of Teaching in Physical Education*, *40*(1), 118–125. https://doi.org/10.1123/jtpe.2019-0138

SHAPE America. (2014). *National standards and grade-level outcomes for K–12 physical education*. Human Kinetics.

SHAPE America. (2017). *National Standards for Initial Physical Education Teacher Education*. https://www.shapeamerica.org/accreditation/upload/National-Standards-for-Initial-Physical-Education-Teacher-Education-2017.pdf

Tsuda, E., Ward, P., Kim, J., He, Y., Sazama, D., & Brian, A. (2021). The tennis common content knowledge measure validation. *European Physical Education Review*, *27*(3), 654–665. https://doi.org/10.1177%2F1356336X20982629

Tsuda, E., Ward, P., Li, Y., Higginson, K., Cho, K., He, Y., & Su, J. (2019). Content knowledge acquisition in physical education: Evidence from knowing and performing by majors and nonmajors. *Journal of Teaching in Physical Education*, *38*(3), 221–232. https://doi.org/10.1123/jtpe.2018-0037

Tsuda, E., Ward, P., Yoshino, S., Ogiwara, T., He, Y., & Ohnishi, Y. (2019). Validity and reliability of a volleyball common content knowledge test for Japanese physical education pre-service teachers. *International Journal of Sport and Health Science, 17*, 178–185. https://doi.org/10.5432/ijshs.201916

Viera, B., & Ferguson, B. (1989). *Teaching volleyball: Steps to success*. Leisure Press.

Ward, P. (2009). Content matters: Knowledge that alters teaching. In L. Housner, M. W. Metzler, P. G. Schempp, & T. J. Templin (Eds.), *Historic traditions and future directions of research on teaching and teacher education in physical education* (pp. 345–356). Fitness Information Technology.

Ward, P., & Ayvazo, S. (2016). Pedagogical content knowledge: Conceptions and findings in physical education. *Journal of Teaching in Physical Education*, *35*(3), 194–207. https://doi.org/10.1123/jtpe.2016-0037

Ward, P., & Lehwald, H. (2018). *Effective physical education content and instruction*. Human Kinetics.