Old Dominion University

# ODU Digital Commons

# Assessing Teacher Pedagogical Knowledge: The Video Assessment of Teacher Knowledge (VATK)

Peter D. Wiens

Jori S. Beck
*Old Dominion University*, jbeck@odu.edu

Christina J. Lunsmann

Follow this and additional works at: https://digitalcommons.odu.edu/teachinglearning_fac_pubs

Part of the Educational Assessment, Evaluation, and Research Commons, and the Educational Technology Commons

## Original Publication Citation

**Assessing Teacher Pedagogical Knowledge: The Video Assessment of Teacher Knowledge (VATK)**

*Peter D. Wiens[1], Jori S. Beck[2], and Christina J. Lunsmann[3]

[1]*University of Nevada, Las Vegas, Las Vegas, NV, USA*
[2]*Old Dominion University, Norfolk, VA, USA*
[3]*University of South Carolina Sumter, Sumter, SC, USA*

Corresponding Author

Peter Wiens, peter.wiens@unlv.edu, 4505 S. Maryland Parkway Box 453005, Las Vegas, NV, 89154-3005, USA

Biography Statements:

Peter D. Wiens, Ph.D., is Assistant Professor of Teacher Education in the Department of Teaching and Learning at the University of Nevada, Las Vegas. His research focuses on the assessment of teacher education outcomes and preservice teacher learning including the last effects of teacher preparation on instructional outcomes.

Jori S. Beck, Ph.D. is an Assistant Professor of Secondary Education in the Department of Teaching and Learning at Old Dominion University. Her research interests focus on clinical teacher preparation broadly with emphases in Third Space partnerships, data literacy for teaching, and equity.

Christina J. Lunsmann, Ph.D., is the Program Coordinator for the University of South Carolina, Sumter's teacher education program through USC Aiken. Her work focuses on teacher preparation with specific attention to democratic teaching practices, constructivism, and mentorship.

**Abstract**

This study examines the creation and implementation of a video-based assessment of teacher knowledge of effective teaching called the Video Assessment of Teacher Knowledge (VATK).  The VATK was developed with the intention of creating an easily administered, standardized assessment that can potentially provide information on teacher candidate knowledge of teaching at important points during their training programs and into the teaching field. The theoretical framework upon which the VATK was developed and the process for item creation are described. Classical test theory and item response theory analysis were conducted to determine the measure's potential for use in future studies. Analyses indicate that after some questions are removed, the measure is sensitive to differences in teachers' pedagogical knowledge.

*Keywords:* video; assessment; teacher preparation; teacher noticing; item response theory, teacher expertise

**Assessing Teacher Pedagogical Knowledge: The Video Assessment of Teacher**

**Knowledge (VATK)**

Many countries are witnessing the rise of performance standards (Cowie and Cooper 2017) and increased accountability for teacher preparation. Recent years have seen increased pressure on teacher preparation programs in the United States—the location of the research reported here—to demonstrate that teacher candidates develop capacity for practicing effective teacher-student interactions that will promote their students' learning. There has been progress in developing standardized, reliable end-of-program assessments such as the edTPA (Darling-Hamond 2006; Pearson Education, Inc. 2018) and tests such as the PRAXIS exams in the United States (Educational Testing Services 2018). However, these assessments do not provide information about growth in teacher candidate knowledge and abilities throughout their programs. Thus, there is a need for theoretically and empirically supported measures that can provide reliable data on teacher candidate growth.

The current study examines the creation and implementation of a video-based assessment of teacher knowledge of effective teaching called the Video Assessment of Teacher Knowledge (VATK). The VATK was developed with the intention of creating an easily administered, standardized assessment that could potentially provide information on teacher candidate knowledge of teaching at important points during their training programs and even into the teaching field. In this paper, we describe the theoretical framework upon which the VATK was developed, the process for creating items, the initial data collected using the instrument, and the psychometric testing of the measure.

**Theoretical Framework**

Decades of research based on observing teachers in the field has provided a strong indication of what is and is not effective in teaching (Brophy and Good 1986; Pianta and Hamre 2009). Built upon this research, traditional measures such as the Framework for

Teaching (Danielson 1996), the Classroom Assessment Scoring System (CLASS; Pianta and Hamre 2009), and the edTPA (Darling-Hammond 2006) provide standardized measures that have been theoretically and empirically supported. While these measures can provide a viable way to assess teaching quality, they rely on observations of practicing teachers. Teacher education programs need standardized measures that can provide information about future teachers' abilities when these individuals are not yet teaching in the field. Therefore, it is necessary to develop measures that gauge skills that relate to teacher candidates' future teaching quality.

### Identifying Effective Teaching is related to Effective Teaching

It is possible to use an individual's knowledge of effective teaching as demonstrated in the ability to recognize effective teaching in videos as an assessment. The 'Learning to Notice' framework created by Van Es and Sherin (2002) describes the importance of teachers' abilities to identify features of complex classroom situations and is based on expertise research that indicates experts see situations differently and in more complex ways than do novices (Feldon 2007; Glaser and Chi 1988). In a study of math teachers, Kersting, Givvin, Sotelo, and Stigler (2010) found an association between the quality of teachers' video analysis and their students' learning. Kaiser and colleagues (2015), in their development of an instrument to measure mathematics teachers' knowledge, concluded that it is possible to capture the way that teachers notice classroom situations. Similarly, Kersting (2008) sought to measure teachers' knowledge of teaching mathematics and discovered a relationship between video-analysis scores and mathematical content knowledge for teaching. Likewise, Wiens, LoCasale-Crouch, Cash and Romo (2020), showed that preservice teachers' ability to identify effective teaching interactions in videos was predictive of their ability to implement those effective teaching interactions in their student teaching experience. Thus, considering how to assess teachers' ability to identify effective interactions is worth pursuing.

### *A Standardized Growth Measure*

Teacher candidates are generally evaluated using different means, including basic skills, content knowledge, and pedagogical knowledge assessment that are required for licensure (Crowe 2010). These assessments are often completed before teacher candidates enter or complete teacher education programs and are not necessarily valuable in determining which candidates will be successful teachers (Darling-Hammond 2010). These question-and-answer assessments are the focus of much teacher knowledge assessment research (Baumert et al. 2010). Since the 1990s, these performance assessments have been developed as 'portfolios that collect evidence of teachers' actual instruction, through videotapes, curriculum plans, and samples of student work and learning' (Darling-Hammond, Newton, and Wei 2013, 180).

Standardized pedagogical knowledge assessments that can be easily administered throughout the teacher preparation program would be useful, as programs would no longer have to wait until the end of the program to conduct performance-based assessments. An assessment to measure growth in addition to a performance-based assessment at the end of the program could allow teacher preparation programs to provide interventions throughout the teacher candidate's educational experience.

One video-based assessment that allows for assessment at multiple points during the teacher preparation program is the Video Assessment of Interactions and Learning (VAIL; Jamil, Sabol, Hamre, and Pianta 2015). VAIL 'directly assesses teachers' skills in detecting effective classroom interactions from video' (Jamil et al. 2015, 411). Jamil and colleagues examined VAIL data from 270 preschool teachers and found that performance on VAIL was related to participants' ability to implement effective teaching interactions (Jamil et al 2015). Additionally, VAIL scores were positively correlated with years of teaching experience. Studies have demonstrated the ability of the VAIL to be reliably administered at multiple

points in a teacher education program (Wiens, Hessberg, LoCasale-Crouch and DeCoster 2013). However, the VAIL requires open-ended responses from participants using substantial resources to reliably code and analyse. Additionally, an analysis of five years of longitudinal data from the VAIL collected in a teacher education program showed that participants put less effort into the assessment at the end of their program than at the beginning of their program (Wiens and Gromlich 2018).

The Video Assessment of Teacher Knowledge (VATK) was developed as a valid alternative to these video-based measures that could assess teacher knowledge and potentially capture growth in this knowledge. The design of a valid and reliable measure that could be easily administered at multiple points in a teacher education program with low-resource requirements could allow institutions to capture changes in knowledge through longitudinal data collection. However, it should be noted that the VATK and other video-based measures created on the learning to notice framework (Van Es and Sherin 2002) are complementary assessments to others currently used in teacher education including portfolios, lesson plans, and teaching evaluations such as the edTPA (Pearson Education, Inc. 2018). Additionally, there is a need for standardized measures that can be used across contexts (Pianta and Hamre 2009). Therefore, a wide variety of assessments are required to understand teacher candidate learning and preparedness for teaching. In the following section we will discuss the theoretical framework for the VATK and the creation and validation of the measure.

**A Vision of Effective Teaching: The InTASC Standards**

It was important to situate the VATK in a widely-accepted vision of effective teaching. For this, we turned to the Council of Chief State School Officers' (CCSSO) InTASC Standards (CCSSO 2011). The InTASC Standards 'describe what effective teaching that leads to improved student achievement looks like' (CCSSO 2011, 3). These standards have been widely accepted by teacher education organizations across the United States. The

InTASC Standards are divided into four categories—(a) The Learner and Learning, (b) Content Knowledge, (c) Instructional Practice, and (d) Professional Responsibility—which are further divided into ten total standards. We focused on the standards that would most easily be observable in videos of authentic classroom teaching as described below. Of the remaining standards, content knowledge and application of content were not a focus of the VATK because we sought to create a non-subject specific assessment. Furthermore, planning for assessment, professional learning and ethical practice, and leadership and collaboration were also not focused on due to the fact that these would be difficult to observe in a video recording of a class.

### *InTASC Standard #1: Learner development*

Educational research has a long history of support for understanding child development and using this information in instructional decision-making (Snowman and McCown 2015). For example, in the 1930s, Vygotsky (1986) put forward that teachers should recognize students' development levels and cater instruction to their zone of proximal development. More recently, research has concluded that teachers need to understand the knowledge their students possess in order to adjust their teaching to the students' needs (Bransford, Brown, and Cocking 2000). Since students construct their new knowledge based on their previous knowledge and experiences (Wheatley 1991), this is a vitally important skill for teachers.

### *InTASC Standard #2: Learning differences*

The United States is a diverse country with a growing minority population. Understanding the cultural and linguistic needs of students is an important aspect of effective teaching. Furthermore, teachers should be able to build on students' cultural backgrounds to construct positive learning environments (Gay 2002). Classrooms that do not support the linguistic needs of students disadvantage those students' learning (Abedi and Herman 2010).

Likewise, students who experience learning based on their cultural background are more likely to show greater learning gains (Dee and Penner 2017).

### InTASC Standard #3: Learning environments

Managing a classroom effectively is a complex undertaking that requires different skills and attention to various elements. For example, teacher-student relationships have been shown to have a tremendous effect on student learning (Hattie 2009). However, these relationships must be carefully developed so that the teacher is viewed as dominant, yet caring (Marzano, Marzano, and Pickering 2003). Other important facets of classroom management include rules and procedures and disciplinary interventions (Marzano et al. 2003). Withitness, or the ability to quickly and accurately identify potential problematic behaviours, has also been shown to be important to creating a positive learning environment (Kounin, 1970). Effective teachers also communicate effectively with students, guardians, and colleagues from the beginning of the school year the expectations for the class (Emmer and Evertson, 2012).

### InTASC Standard #6: Assessment

In the field of teacher preparation, the construct of assessment literacy (Impara, Plake, and Fager 1993) has dominated discussions regarding what new and veteran teachers need to know and be able to do regarding assessing student learning. Specifically, assessment literacy focuses on multiple assessment and their outcomes including their purpose, achievement targets, methods, outcomes, and bias and distortion (Stiggins 1995). However, the broader construct of data literacy (Gummer and Mandinach 2015) is beginning to take hold. Data literacy focuses on the teacher's ability to transform many different types of data including assessment, school climate, behavioural, cross-sectional, and longitudinal into actionable instructional knowledge. It is important for teachers to understand both assessment and data literacy to accurately assess student learning and drive their instruction.

***InTASC Standard #8: Instructional strategies***

Arends (2015) identified six different types of instructional strategies steeped in research on teaching and learning. He divided these into teacher-centred (presenting and explaining and direct instruction) and student-centred approaches (concept and inquiry-based teaching, cooperative learning, problem-based learning, and classroom discussion). Cooperative learning can be a powerful instructional activity that can have a positive impact on student achievement, interpersonal relationships, and attitudes toward learning (Marzano et al. 2003). New teachers must develop a repertoire of instructional strategies to implement according to student needs and the content at hand (Arends 2015).

These InTASC standards are specifically designed for teacher candidates, while traditional assessments of teacher effectiveness are designed to evaluate practicing teachers. By tailoring the VATK to these standards, teacher preparation programs will have a sense of their teacher candidates' understanding of effective teaching. However, based on the literature presented previously regarding the differences in experts and novices (Feldon 2007; Glaser and Chi 1988) as well as the narrow focus of beginning teachers (Conway and Clark 2003; Guillaume and Rudney 1993) we propose a single-factor structure as a good starting point for examination of teacher pedagogical knowledge as measured by the VATK.

Teacher knowledge is generally considered as multidimensional and sometimes identified as content knowledge, pedagogical content knowledge, and general pedagogical knowledge (GPK) based on Shulman's (1987) seminal work. The VATK is built on a developing line of empirical research examining general content knowledge (Depaepe and König 2018; König, Blömke, Paine, Schmidt, and Hsieh 2011; Lauermann and König 2016; Voss, Kunter, and Baumert 2011). This research shows that teacher knowledge can be examined using the concept of GPK and it also demonstrates that GPK can be measured longitudinally in teacher preparation programs and indicate a valid measurement of GPK can

be understood through a unidimensional framework. Wong, Chong, and Choy (2008) examined preservice teacher perceptions of GPK in a Singapore teacher preparation program and discovered that elementary preservice teachers felt that they developed more GPK than did their secondary peers. König, Ligtvoet, Klemenz, and Rothland (2017), in a study of 1347 student teachers in Germany and Austria that examined longitudinal change in GPK, found that knowledge growth was related to certain aspects of the teacher training programs. While there is sufficient research to support measuring GPK, the field still requires additional research and measures to better understand teacher knowledge.

## Methods

### Procedures

The procedures used in this study were designed to follow established instrument creation procedures based on those outlined by Lester, Inman, and Bishop (2014). Lester and colleagues suggest a four-step process in assessment creation for an initial pilot study of a new measure. In this study, we report on the procedures for the four steps in this process: 1) theory or taxonomy selection, 2) item pool creation, 3) content validation, and 4) pilot study (Lester et al., 2014). The selection of a theory or taxonomy was described in the previous framework section. In the following section, we will describe the creation of an item pool, initial content validation, and our pilot study.

### *Initial item creation*

We selected videos of practicing teachers based on their ability to demonstrate the selected teaching standards. Videos were taken in real-world classrooms at various grade levels and content areas to be representative of a typical, quality classroom environment. Specifically, the second author chose videos from a previous project in which teachers were video recorded for training purposes for a teacher evaluation system used by the State. Participants had signed media releases that transferred their rights to the videos to the

university. As the purpose of the VATK is to assess the knowledge of teachers across grade levels and disciplines, we selected four teachers from different grade levels and subject areas. These included two elementary classrooms (grades two and four), one middle school classroom (science), and one high school classroom (English).

A team of four doctoral students with experience teaching at the various grade levels was recruited to create an initial round of questions. Each doctoral student was assigned one teacher with multiple recorded classes. From the collections of videos from each teacher's classroom, the doctoral students selected short video clips (3-5 minutes) and were asked to develop twenty multiple-choice questions spread across the five selected InTASC Standards (CCSSO 2011). As Haladyna (2004) pointed out, a multiple-choice assessment can be 'efficient and provides a useful summary of student learning or knowledge and cognitive skills' (67). Haladyna developed a list of 31 'General Item-Writing Guidelines' (99-100) that informed our approach to writing good multiple-choice questions.

All seventy-nine questions (see Table 1) were brought to the research team and each question was analysed by at least three individuals who did not write the question. Questions were examined for clarity, difficulty, and validity. The lead researchers then examined the research team feedback and the questions and created an initial version of the assessment that consisted of all four classrooms and a total of 37 questions. This process involved the two lead researchers examining each question for alignment to standards, clarity, and readability.

### Building content validity

The process of developing content followed well-established procedures for assessing validity (Thorndike and Thorndike-Christ 2010). The first version of the VATK was created in the online platform Qualtrics (2018), which included the video clips of the classrooms followed by the multiple-choice questions. Links to the initial VATK were sent to four expert reviewers for comments, and three provided feedback. The expert reviewers examined each

of the questions and provided written feedback through the online survey asking them to provide feedback on each question, the overall measure, and missing items. The experts were all current or retired tenured teacher education professors from three different universities in the United States. Two of the experts had teacher preparation experience at large public universities, while the third had experience in teacher preparation at a smaller, private university (see Table 2).

The expert reviewers' comments were analysed by the lead researchers, and these were used to refine the questions for clarity. At this time, eight questions were eliminated which resulted in a 29-question version of the VATK. The high school English classroom was also eliminated in an effort to streamline the assessment so that participants would have to watch fewer video clips and, therefore, shorten the amount of time it takes to complete the VATK. This became the first version of the VATK that was used for data collection. The number of questions by classroom and InTASC Standard are shown in Table 3.

The new 29-item version of the VATK was then administered to four different doctoral students who have three or more years of experience as K-12 classroom teachers. These individuals were asked to answer each question on the VATK and provide a detailed written explanation of why they answered the question the way they did. Analysis of the doctoral students' answers to the questions as well as their explanations of why they answered the question was used in another round of revisions to the questions for clarity. Minor mistakes in the measure as well as more substantial changes to questions and answer options based on the clarity of the questions, the difficulty of the questions, and the clarity of the answers were completed, and a final version was created for distribution to a group of teachers.

***Data collection***

In order to establish baseline data on the VATK, we recruited practicing teachers in two ways. First, we recruited teachers at a local, private PreK-8 school. These teachers were recruited through the school's administrative team and offered a $25 gift card for participation. To protect teachers' privacy, we collected no demographic data on these individuals so that tracing their data was purposefully extremely difficult. The second group of teachers was recruited through a university-based research subject pool. These teachers were enrolled in university classes in a college of education that required students to complete research credits or alternative assignments. Participants from this group were awarded research credit in their classes for participation in this study. Third, students in introduction to education courses completed the VATK in the first two weeks of their courses. Data were collected through a Qualtrics-hosted (2018) survey over a twelve-month period. The first group of in-service teachers completed the VATK in the summer months, followed by the novice teacher candidates at the beginning of the following fall and spring semesters. Finally, teachers enrolled in the university courses completed the VATK in the last four weeks of the fall semester.

**Participants**

The total number of participants for this study included 153 teacher candidates and in-service teachers. There were 40 practicing teachers. Of the sample, 22 were teachers in a PreK-8 private school. Due to confidentiality concerns, we were unable to ask these teachers their years of experience or teaching position as these would have identified them. Additionally, 18 participants were practicing teachers enrolled in a graduate-level course in the college of education at a large, public university situated in the Southwestern United States. These participants received research participation credit in their course for completing the VATK. This group of practicing teachers had an average of 2.76 years of teaching experience including the current year. The group of teacher candidates consisted of

undergraduate university students and master's students enrolled in an introduction to education course. These students were required to complete the VATK as well as a survey in the first two weeks of their course, but given the option of allowing their responses to be used for research purposes. Data for this study were taken from one semester of introductory courses and included 102 participants. A final group of participants from the university course with a research requirement (n=12) were not used in this study due to contradictory responses in their survey regarding whether they were currently a full-time teacher or a teacher candidate.

**Instrument**

The VATK is a 29-item, multiple-choice assessment of teacher knowledge of effective teaching techniques. The VATK consists of brief clips of videos of elementary and middle school classrooms. Participants watch six, three-to-five-minute video clips (two each of second grade, fourth grade, and middle school science). Questions follow each video and are addressed to the specific video immediately preceding the question. Each question has four answer options. Some questions are 'select all' and some are limited to one answer selection. Sample questions for each construct can be seen in Table 4.

Participants first watch the videos (which may be re-watched as many times as needed). Each of the six videos is between three and five minutes long. Then participants select the correct -answer choice depending on the question. Based on the InTASC Standards (CCSSO 2011), the VATK assesses a participant's understanding of general teaching strategies. Each participant earns a score based on a 29-point scale, earning one point for each question. Each question must be answered completely correctly to earn a point. No partial credit was possible for any of the questions.

**Analysis**

Analysis of the VATK occurred in two phases. First, we conducted classical item analysis, and reliability of the scores—internal consistency—was estimated (Crocker and Algina 1986). Next, we employed item response theory exploratory analyses (IRT) including Rasch Models for dichotomous responses to the VATK (de Ayala 2009). Finally, analysis of variance (ANOVA) was calculated to understand group differences between practicing teachers and entry-level teacher candidates. Classical item analysis and ANOVA analysis were conducted in SPSS version 24 (IBM Analytics n.d.). IRT analysis was conducted using the jMetrik (Meyer 2014) analysis software.

Prior to analysis, all questions were coded for correctness. All VATK questions were scored as correct (1) or incorrect (0). Missing responses were coded as incorrect (0). Therefore, there were no missing data in the dataset.

## Results

### Classical Item Analysis

Classical item statistics including item difficulties (item mean values), and discrimination (corrected item total correlations) were computed for the VATK. Analysis of item mean values indicated a wide range of difficulty from $M_5 = .026$ to $M_{20} = .863$. Therefore, some questions were answered correctly by under 3% of participants while others were answered correctly by more than 85% of participants.

Corrected item-total correlation also ranged widely from item to item with a range of .025 to .419. Three items with item-total correlation values lower than .1 were flagged for further analysis. The estimate of internal consistency among the 29 items was $\alpha=.767$. A final score was calculated by summing up the corrected responses. The score distribution of the total score ranged from 1 to 25 on the 29-point scale. The mean score was 11.183 with a standard deviation of 4.576. There did not appear to be any floor or ceiling effects.

### Item Reponses Theory Analysis

IRT refers to a family of analyses that 'uses latent characterizations of individuals and items as predictors of observed responses' (de Ayala 2009, p. 4). In this study, we conducted a dichotomous analysis for instrument validation (de Ayala 2010). While classical test theory bases the analysis of items on the sample population, IRT assumes participant ability as a latent trait called theta ($\Theta$). In this study, we focused on the Rasch Model of IRT which 'uses a logistic function to represent the probability that a randomly selected examinee with ability level $\Theta$ correctly answers an item' (Meyer 2014, 83).

The jMetrik uses the joint maximum likelihood estimate (JMLE) parameters for Rasch Model (see Meyer 2014). Using the JMLE, we conducted item fit analyses to examine the INFIT and OUTFIT for each of the 29 items and is illustrated in Table 5. A common metric for examining these scores is that items that score between .5 and 1.5 are considered acceptable (de Ayala 2010). Values closer to 1 may be required for a high stakes test (Meyer 2014) but were an adequate benchmark for our analyses. From our Rasch analysis, we found that item difficulty ranged from -2.86 to 3.42 with a mean of 0. INFIT mean square statistics were between .85 and 1.24 with a mean value of .998. Meanwhile, mean square OUTFIT statistics ranged from .60 to 1.81 with a mean of 1.041. Using all items from the VATK, person ability estimates ($\Theta$) range from -4.07 to 2.88 with a mean of -.69 and standard deviation of 1.05. The distribution of examinees is mostly equal to the distribution of items as shown in Figure 1; however, there appears to be one item that is more difficult than participant ability and some participants whose ability is less than the easiest question. Scale quality statistics can be seen in Table 6 and show a person reliability of .78 and person separation index of 1.91.

Based on our classical item analysis and our IRT, we determined that we would remove four items (numbers 1, 2, 8, and 27); however, we maintained construct validity as

shown in Table 7. With the four items removed, the VATK has 25 items with the following descriptive statistics M=10.84, SD=4.43 and a reliability coefficient of α=.772.

**Group Difference Analysis**

In our final analysis, we examined if practicing teachers and teacher candidates at the beginning of an introductory course would score differently on the VATK. Using the 25-question version of the VATK we conducted an ANOVA to test the group mean difference in their scores. The teachers (M=12.26, SD=5.766) had a higher mean score than the prospective teachers (M=10.24, SD=3.73). This difference was statistically significant F=5.99, p=.016 indicating that the difference was not attributable to sampling error.

## Discussion

As a complement to current and future assessment strategies, the VATK shows promise as having validity and reliability evidence to support its use in teacher education. Data presented in this paper indicate that the VATK could be used efficiently and effectively in teacher education; it is possible to create instruments that are not cumbersome to administer but differentiate between novice and expert teachers.

The VATK has a strong theoretical base and data presented in this paper demonstrate the ability of the VATK to be implemented in a teacher education setting as well as with in-service teachers to collect information about participants' knowledge of teaching pedagogy. Given the need to develop early markers of teacher knowledge (Pianta and Hamre 2009), the VATK has potential to provide valuable information to teacher preparation programs and the empirical literature by tracking pedagogical knowledge growth longitudinally. Even though the data collected in this study were cross-sectional, they demonstrate the practicality of using the VATK across different time points in the teaching career.

One goal of the design of the VATK was to create an instrument that was valid, reliable, and economical. Previous video-based measures (Jamil et al. 2015; Kersting 2008;

König et al. 2014) have demonstrated strong validity and reliability, but they have drawbacks when it comes to how easily and economically they are administered. The VATK was administered economically, but it also shows the ability to differentiate between expert and novice teachers. Based on expertise research (Feldon 2007), experts observe complex situations in different ways than novices do, so it is important that the VATK can detect these differences.

The noticing framework (Van Es and Sherin 2002) assumes that teachers and teacher candidates can be trained to improve their noticing ability. Likewise, there is a connection between noticing ability and teaching ability (Kersting 2008). In our sample, experienced teachers were better able to answer pedagogical questions related to real-life classrooms. Future research will need to follow the same teacher candidates through their teacher preparation programs to understand how these skills develop over time in individuals. The VATK shows promise as a measure that can detect these differences in a standardized manner to supplement traditional, observational measures of teacher effectiveness.

## Limitations

In this data collection effort, we had two important limitations to what data we could collect on our participants. We were unable to collect demographic information about the participants due to privacy concerns with our small sample of teachers, and, therefore, we could not test for differential item functioning—something that will need to be included in future analyses. While we assume the practicing teachers in our sample are 'experts' based on their teaching experience, but we have no way to verify their actual teaching ability. In future studies it would be helpful to use observational measures such as the Classroom Assessment Scoring System (Pianta and Hamre 2009) or the Framework for Teaching (Danielson and McGreal 2000) to identify expert teachers and exam the correlations of the VATK with observed teaching performance. We can only base this assumption on their years of

experience as classroom teachers. Future studies should use a measure of teaching

effectiveness to understand the relationship between performance on the VATK and ability to

teach well.

      We should also point at that the videos were recorded in classrooms in the U.S. The

videos and pedagogical concepts in the VATK were intended to be broadly applicable across

many contexts. We attempted to avoid connections to U.S. society or culture; however, future

research would need to be conducted to know if the VATK is a valid measure in other

countries.

### Implications for Future Research, Policy, and Practice

      The VATK is in the beginning stages of research, and there are many directions that

future research should take. Because of the robust body of research related to assessing

teacher effectiveness, it seems pertinent to study the correlation, if any, between teachers'

scores on those assessments and the VATK. Also, while the pilot study allowed us to

examine the differences in responses of practicing and prospective teachers, it did not allow

us to observe participants' scores at multiple stages in their development. A longitudinal

study that observes growth over time will demonstrate the instrument's effectiveness.

      Currently, many countries have policies that require prospective and practicing

teachers to be assessed by various subjective measures. The VATK could provide a more

objective measure for administrators in determining teacher readiness. Coupling VATK with

other measures may provide another level of evaluative certainty to teachers and

administrators.

## References

Abedi, J. and J. Herman (2010). Assessing English language learners' opportunity to learn

    mathematics: Issues and limitations. *Teachers College Record, 112*(3), 723-746.

Arends, R. I. (2015). *Learning to teach.* 10th ed. New York: McGraw-Hill.

Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, … Y-M. Tsai, (2010).

    Teachers' mathematical knowledge, cognitive activation in the classroom, and student

    progress. *American Educational Research Journal, 47(*1), 133-180.

Bransford, J. D., A. L. Brown, and R. R. Cocking, (2000). *How people learn: Mind, brain,*

    *experience, and school.* Washington, DC: National Research Council.

Brophy, J. E., and T. L. Good, (1986). Teacher behaviour and student achievement. In M. C.

    Wittrock (Ed.), *Handbook of research on teaching,* (pp. 328-375). New York:

    Macmillan.

Conway, P. F. and C. M. Clark, (2003). The journey inward and outward: A re-examination

    of Fuller's concerns-based model of teacher development. *Teaching and Teacher*

    *Education, 19,* 465-482. doi:10.1016/S0742-051X(03)00046-5

Council of Chief State School Officers. (2011). *InTASC model core teaching standards: A*

    *resource for state dialogue.* Washington, DC: Author.

Cowie, B., and B. Cooper, (2017). Exploring the challenge of developing student teacher data

    literacy. *Assessment in Education: Principles, Policy and Practice, 24*(2), 147-163.

    doi:10.1080/0969594X.2016.1225668

Crocker, L. and J. Algina, (1986). *Introduction to classical and modern test theory.* Orlando:

    Holt, Rinehart and Winston.

Crowe, E. (2010). *Measuring what matters: A stronger accountability model for teacher*

    *education.* Washington, D.C.: Center for American Progress.

Danielson, C. (1996). *Enhancing professional development: A framework for teaching.*

    Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C., & T.L. McGreal. (2000). *Teacher evaluation to enhance professional*

    *practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple

    methods for assessing program outcomes. *Journal of Teacher Education, 57*, 120-138.

    https://doi.org/10.1177/0022487102053004002

Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance*

    *assessments can measure and improve teaching.* Washington, D.C.: Center for

    American Progress.

Darling-Hammond, L., S. P. Newton, and R. C. Wei, (2013). Developing and assessing

    beginning teacher effectiveness: The potential of performance

    assessments. *Educational Assessment, Evaluation and Accountability, 25*(3), 179-204.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York:

    Guilford Press.

de Ayala, R. J. (2010). Item Response Theory. In G. R. Hancock, R. O. Mueller, and L. M.

    Stapleton. *The reviewer's guide to quantitative methods in the social sciences* (pp.

    155-172). New York: Routledge.

Dee, T. S. and E. K. Penner, (2017). The causal effects of cultural relevance: Evidence from

    an ethnic studies curriculum. *American Education Research Journal, 54*(1), 127-166.

    doi:10.3102/0002831216677002

Depaepe, F., and J. König, (2018). General pedagogical knowledge, self-efficacy and

    instructional practice: Disentangling their relationship in pre-service teacher

    education. *Teaching and Teacher Education, 69*, 177-190.

    https://doi.org/10.1016/j.tate.2017.10.003

Educational Testing Service. (2018). *Praxis.* Retrieved from https://www.ets.org/praxis/

Emmer, E. T. and C. M. Evertson. (2012). *Classroom management for middle and high school teachers.* New York, NY: Pearson.

Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review, 19*, 91-110. doi:10.1007/s10648-006-9009-0.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*(2), 106-116.

Glaser, R. and M. T. H. Chi, (1988). Overview. In M.T.H. Chi, R. Glaser, and M.J. Farr (Eds.). *The nature of expertise* (pp. xv-xxviii). Mahwah, NJ: Lawrence Erlbaum Associates.

Guillaume, A. M. and G. L. Rudney, (1993). Student teachers' growth toward independence: An analysis of their changing concerns. *Teaching and Teacher Education, 9,* 65-80.

Gummer, E., and E. B. Mandinach, (2015). Building a conceptual framework for data literacy. *Teachers College Record, 117,* 1-22.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. 3rd ed. Mahwah: Lawrence Erlbaum Associates.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

IBM Analytics. (n.d.) *ISBM SPSS software.* Retrieved from https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software

Impara, J. C., B. S. Flake, and J. J. Fager. J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership, 3,* 510-521.

Jamil, F., T. Sabol, B. Hamre, and R. Pianta, (2015). Assessing teachers' skills in detecting
        and identifying effective interactions in the classroom. *The Elementary School
        Journal, 115*(3), 407-432.

Kaiser, G., A. Busse, J. Hoth, J. König, and S. Blomeke (2015). About the complexities of
        video-based assessments: Theoretical and methodological approaches to overcoming
        shortcomings of research on teachers' competence. *International Journal of Science
        and Math Education, 13*, 369-387. doi:10.1007/s10763-015-9616-7

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts
        to measure teachers' knowledge of teaching mathematics. *Educational and
        Psychological Measurement, 68*(5), 845-861. doi:10.1177/0013164407313369

Kersting, N. B., K. B. Givvin, F. L. Sotelo, and J. W. Stigler, (2010). Teachers' analyses of
        classroom video predict student learning of mathematics: Further explorations of a
        novel measure of teacher knowledge. *Journal of Teacher Education, 61*(1-2), 172-
        181. doi:10.1177/0022487109347875.

König, J., S. Blömeke, L. Paine, W. H. Schmidt, and F. J. Hsieh, (2011). General pedagogical
        knowledge of future middle school teachers: On the complex ecology of teacher
        education in the United States, Germany, and Taiwan. Journal of Teacher Education,
        62(2), 188–201. https://doi.org/10.1177/0022487110388664

König, J., S. Blomeke, P. Klein, U. Suhl, A. Busse, and G. Kaiser, (2014). Is teachers'
        general pedagogical knowledge a premise for noticing and interpreting classroom
        situations? A video-based assessment approach. *Teaching and Teacher Education, 38*,
        76-88.

König, J., R. Ligtvoet, S. Klemenz, and M. Rothland, (2017). Effects of opportunities to learn
        in teacher preparation on future teachers' general pedagogical knowledge: Analyzing

program characteristics and outcomes. *Studies in Educational Evaluation, 53*, 122–

133. https://doi.org/10.1016/j.stueduc.2017.03.001

Kounin, J. S. (1970). *Discipline and group management in classrooms.* New York, NY: Holt,

Rinehart, & Winston.

Lauermann, F., and J. König (2016). Teachers' professional competence and wellbeing:

Understanding the links between general pedagogical knowledge, self-efficacy and

burnout. Learning and Instruction, 45, 9–19. DOI: 10.1016/j.learninstruc.2016.06.006

Lester, P. E., D. Inman, and L. K. Bishop (2014). *Handbook of tests and measurement in

education and the social sciences* (3rd ed.). Lanham: Rowman and Littlefield.

Marzano, R. J., J. S. Marzano, and D. Pickering, (2003). *Classroom management that works:

Research-based strategies for every teacher*. Alexandria, VA: Association for

Supervision and Curriculum Development.

Meyer, J. P. (2014). *Applied measurement with JMetrik.* New York, NY: Routledge.

Pearson Education, Inc. (2018). *edTPA*. Retrieved from https://www.qualtrics.com

Pianta, R. C. and B. K. Hamre (2009). Conceptualization, measurement, and improvement of

classroom processes: Standardized observation can leverage capacity. *Educational

Researcher, 38*(2), 109-119. doi:10.3102/0013189X09332374

Qualtrics. (2018). *Home page*. Retrieved from https://www.qualtrics.com

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard

Educational Review*, *57*, 1–22.

Snowman, J. and R. McCown, (2015). *Psychology applied to teaching,* 14th ed. Stamford:

Cengage.

Stiggins, R. J. (1995). Assessment literacy for the 21st Century. *Phi Delta Kappan, 77*, 238-

245.

Thorndike, R. M. and T. M. Thorndike-Christ, (2010). *Measurement and evaluation in psychology and evaluation*. 8th ed. New York: Pearson.

Van Es, E. A., and M. G. Sherin, (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*(4), 571-596.

Voss, T., M. Kunter, and J. Baumert, (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology, 103*(4), 952–969. DOI: 10.1037/a0025125

Vygotsky, L. S. (1986). *Thought and language.* (A. Kozulin, Trans.) Cambridge: MIT Press.

Wheatley, G. H. (1991). Constructivist perspectives on science and mathematics learning. *Science Education, 75*(1), 9-21.

Wiens, P. D., Hessberg, K., LoCasale-Crouch, J., & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers' knowledge related to effective teaching. *Teaching and Teacher Education, 33,* 24-33. doi: 10.1016/j.tate.2013.01.010

Wiens, P. D. & Gromlich, M. D., (2018).  Five years of video-based assessment data: Lessons from a teacher education program. *Research & Practice in Assessment, 13,* 51-61.

Wiens, P. D., LoCasale-Crouch, J., Cash, A., & Romo F. (2020). Preservice teachers' ability to detect and implement effective teaching interactions: Examining the relationship and predictors. *Journal of Teacher Education*. https://doi.org/10.1177/0022487120910692

Wong, A. F. L., S. Chong, and D. Choy, (2008). A comparison of perceptions of knowledge and skills held by primary and secondary teachers: From the entry to exit of their

preservice programme. *Australian Journal of Teacher Education, 33*(3).

https://doi.org/10.14221/ajte.2008v33n3.6

**Tables**

Table 1
*InTASC Standards and Corresponding Number of Questions at the Beginning of Instrument Development*

| InTASC Standard | Grade 2 | Grade 4 | Middle School Science | High School English | Total |
|---|---|---|---|---|---|
| Standard 1: Learner Development | 2 | 2 | 2 | 2 | 8 |
| Standard 2: Learning Differences | 4 | 5 | 4 | 5 | 18 |
| Standard 3: Learning Environments | 4 | 4 | 5 | 4 | 17 |
| Standard 6: Assessment | 2 | 4 | 6 | 5 | 17 |
| Standard 8: Instructional Strategies | 2 | 5 | 8 | 4 | 19 |
| Total | 14 | 20 | 25 | 20 | 79 |

Table 2
Expert Qualifications for Review of Initial VATK

| | Years of Teacher Education Experience | Setting |
|---|---|---|
| Expert 1 | 36 | State University in the Eastern U.S.A. |
| Expert 2 | 6 | Private University in the Eastern U.S.A. |
| Expert 3 | 20 | State University in the Western U.S.A. |

Table 3

*InTASC Standards and Corresponding Number of Questions in the Tested VAT-K*

| InTASC Standard | Grade 2 | Grade 4 | Middle School Science | Total |
|---|---|---|---|---|
| Standard 1: Learner Development | 1 | 1 | 1 | 3 |
| Standard 2: Learning Differences | 3 | 2 | 2 | 7 |
| Standard 3: Learning Environments | 3 | 1 | 3 | 7 |
| Standard 6: Assessment | 2 | 1 | 4 | 7 |
| Standard 8: Instructional Strategies | 0 | 2 | 3 | 5 |
| Total | 9 | 7 | 13 | 29 |

Table 4

Sample VATK Questions for Each InTASC Construct

| InTASC Standard | Question | Answer Choices |
|---|---|---|
| INTASC 1 | How does this clip demonstrate student-centered learning? Select all that apply. | • The students are moving around the classroom.<br>• The students are engaged in experiential learning.<br>• The students are collaborating.<br>• The students are using candy to demonstrate a new concept. |
| INTASC 2 | What are some methods that this teacher could use to make the content more accessible to English learners? *Select all that apply.* | • He could create a visual representation of each new academic vocabulary word.<br>• He could allow students to write their responses in their native language.<br>• He could allow students to draw depictions of the rocks instead of writing their observations.<br>• He could include a list of root words for the new academic vocabulary. |
| INTASC 3 | Which of the following is an interactive technology that could help enhance this lesson? | • A PowerPoint presentation on the different types of rocks.<br>• A video that demonstrates the rock cycle. |

|  |  |  |  |
|---|---|---|---|

| INTASC 6 | Throughout both videos, there is one student who clearly understands the lesson and wants to answer each question the teacher asks. What is the best way to differentiate the lesson for that student? | • A Google Doc where students can collaborate by sharing their observations.<br>• A podcast about fossils.<br>• Provide her with additional rocks that may require more nuanced observation and more thoughtful comparison.<br>• Ask her to help her classmates who are struggling.<br>• Allow her to start the following day's assignment so that she can get ahead of her classmates.<br>• Ask her to help the teacher grade other students' work so she can see what mistakes others are making. |
| INTASC 8 | In what way could the teacher ensure that he adjusts instruction based on student responses to the pressure demonstration? | • The teacher could have students write their hypotheses on white boards to determine if there are misconceptions.<br>• The teacher could call on specific students and ask them questions about the demonstration.<br>• The teacher could continue with the lesson as planned without using formative assessment.<br>• The teacher could conduct another lab to see if students' misconceptions are corrected. |

Table 5

Rasch Analysis Results

| Analysis | Range | Mean |
|---|---|---|
| Item Difficulty | -2.86 to 3.42 | .000 |
| INFIT | .85 to 1.24 | .998 |
| OUTFIT | .60 to 1.81 | 1.041 |
| Ability Estimates | -4.07 to 2.88 | -.690 |

Table 6
*Scale Quality Statistics*

| Statistic | Items | Persons |
|---|---|---|
| Observed Variance | 1.96 | 1.09 |
| Observed Std. Dev. | 1.40 | 1.04 |
| Mean Square Error | .05 | .23 |
| Root MSE | .23 | .48 |
| Adjusted Variance | 1.91 | .85 |
| Adjusted Std. Dev. | 1.38 | .92 |
| Separation Index | 5.99 | 1.91 |
| Number of Strata | 8.33 | 2.88 |
| Reliability | .97 | .78 |

Table 7
*InTASC Standards and Corresponding Number of Questions in the FINAL VAT-K*

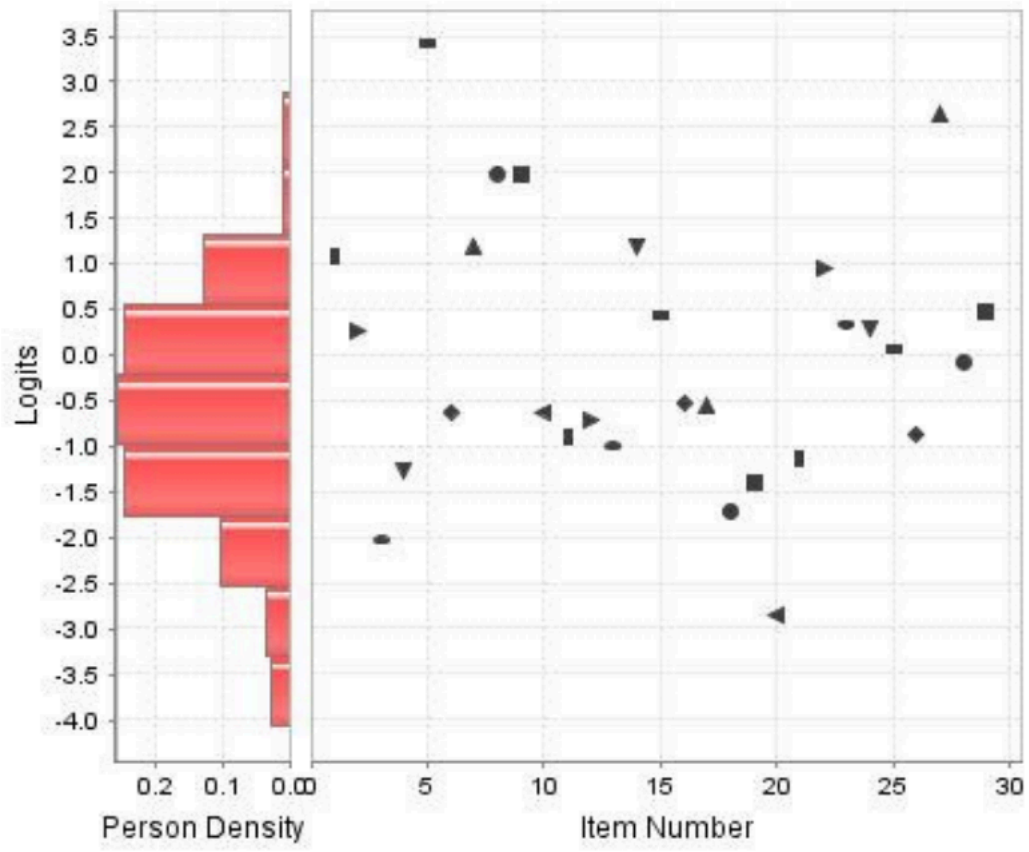| InTASC Standard | Grade 2 | Grade 4 | Middle School Science | Total |
|---|---|---|---|---|
| Standard 1: Learner Development | 1 | 1 | 1 | 3 |
| Standard 2: Learning Differences | 3 | 2 | 1 | 6 |
| Standard 3: Learning Environments | 3 | 1 | 1 | 5 |
| Standard 6: Assessment | 2 | 0 | 4 | 6 |
| Standard 8: Instructional Strategies | 0 | 2 | 3 | 5 |
| Total | 9 | 6 | 10 | 25 |

**Figures**

*Figure 1.* Item Map for VATK