Abstract

Teacher education programs are under considerable pressure to evaluate their effectiveness in training new teachers. Over the last several decades there have been repeated calls for more systematic research on preservice teacher preparation programs. One institution has heeded this call, using the Video Assessment of Interactions and Learning (VAIL: Jamil, Sabol, Hamre, & Pianta, 2015) to annually assess teacher noticing of teacher-student interactions in preservice teachers. While there was no significant difference in first-test and last-test VAIL scores, VAIL scores were shown to be related to final college GPA. Additionally, there was a difference in student effort on the VAIL assessment, as participants provided fewer responses at the end of their program than at the beginning. These data show that assessments can be overused with regard to preservice teacher education programs. Future VAIL implementation should consider assessment fatigue when designing preservice teacher evaluation.

AUTHORS

Peter D. Wiens, Ph.D.
*University of Nevada,
Las Vegas*

Matthew D. Gromlich, Ed.D.
*University of Nevada,
Las Vegas*

Five Years of Video-Based Assessment Data: Lessons from a Teacher Education Program

Teacher education programs are under considerable pressure to evaluate their effectiveness in training new teachers. These pressures come from a variety of sources including state and federal governments, accreditation agencies, media, and potential preservice teachers (Feuer, Floden, Chudowsky, & Ahn, 2013). The fractured nature of the United States education system has led to different approaches across various institutions, as each institution must answer to its own unique set of stakeholders (Feuer et al., 2013). The Council for the Accreditation of Education Programs (CAEP: Council for the Accreditation of Education Programs, 2013), the largest national teacher education accreditation agency in the United States, includes standards requiring programs to show evidence that graduates are ready to be effective teachers. Yet, there remains a need in the field of teacher education to collect data that allows policymakers, researchers, and educators to better understand preservice teachers and the nature of their learning.

There have been repeated calls over the past decades for more systematic research on teacher education (Grossman & McDonald, 2008; Worrell et al., 2014; Zeichner, 2005). Reports issued by the National Academy of Education (Feuer et al., 2013) and an American Psychological Association Task Force (Worrell et al., 2014) provide guidance for the evaluation of teacher education programs. Both reports speak to the difficulty of effectively evaluating such a complex endeavor as training new teachers. Feuer and colleagues (2013) speak to the need for programs to use their core principles when designing evaluation systems, and Worrell and colleagues (2014) state:

CORRESPONDENCE

Email
peter.wiens@unlv.edu.

> The data and methods required to evaluate the effectiveness of teacher education programs ought to be informed by well-established scientific methods that have evolved in the science of psychology, which at its core addresses the measurement of behavior. (p. 2)

Pianta and Hamre (2009) call for studies that identify early markers of teacher quality through standardized measures. Using such measures, links can be drawn between certain programmatic relationships and quality teaching interactions (Pianta & Hamre, 2009).

Teacher education programs use a variety of measures to provide programmatic data in order to address the requirements of various stakeholders (Feuer et al., 2013). One teacher education program made the decision to add to its assessment portfolio by adopting an empirically and theoretically supported video-based assessment of teachers' ability to identify effective teaching interactions. The Video Assessment of Interactions and Learning (VAIL: Jamil, Sabol, Hamre, & Pianta, 2015) was administered to all preservice teachers each year they were in the teacher education program. After five continuous years of data collection, this paper examines what lessons can be learned about the assessment and its usefulness for teacher education evaluation. This data-gathering effort is unique in the teacher education field and examining the data provided may inform other teacher education programs in designing their own assessment frameworks.

Video as an Assessment Tool

Teacher education programs around the world have adopted the use of video for training purposes (Christ, Arya, & Chiu, 2017; Gaudin & Chaliès, 2015). However, assessing preservice teachers' ability to examine videos of real-world classrooms is less frequently cited in the literature. The VAIL, built upon an empirical and theoretical framework that includes teacher noticing and teacher-student interactions, uses video analysis as a means of assessing preservice teachers' knowledge of teaching interactions. The noticing framework is first credited to Goodwin (1994) who wrote, "Professional vision is perspectival, lodged within specific social entities, and unevenly allocated" (p. 626). Van Es and Sherin (2002) further refined the noticing framework in the context of teacher education. They contend that there are three key components of noticing:

(a) identifying what is important or noteworthy about a classroom situation;
(b) making connections between the specifics of classroom interactions and
the broader principles of teaching and learning they represent; and (c) using
what one knows about the context to reason about classroom interactions.
(van Es & Sherin, 2002, p. 573)

Noticing includes observing a situation, making interpretations, and then making a decision on what has been observed (Kaiser, Busse, Hoth, König, & Blömeke, 2015).

A key component of effective noticing is that experts in a field notice different things when observing a situation than do novices—a concept that has been developed and supported in cognition research (Feldon, 2007). Glasser and Chi (1988) developed a list of expert characteristics that includes perceiving problems at a deeper level and spending a larger time analyzing problems. Expert teachers examine information in different ways than do novice teachers (Bransford, Brown, & Cocking, 1999). When examining still photographs, Carter, Cushing, Sabers, Stein, and Berliner (1998) found that experts were more cautious in their observations and were more sensitive to the sequence of events in the classroom, whereas in the case of the novices, "the schema they brought to these visual information processing tasks did not seem as richly developed as experts" (p. 31). Using the expertise framework, the examination of videos can be used as an assessment to differentiate between novices and more expert teachers.

Yet, there remains a need in the field of teacher education to collect data that allows policy makers, researchers, and educators to better understand preservice teachers and the nature of their learning.

Video as an assessment tool has promise because the ability to effectively notice in videos of teaching has been shown to correlate with the ability to teach effectively (Kersting, Givvin, Sotelo, & Stigler, 2010; Santagata & Yeh, 2014). Specifically, the VAIL has been associated with observed teaching quality in both in-service (Hamre, et al., 2012) and preservice teachers (Wiens, 2014). The VAIL is an assessment conducted online where participants watch short videos (2–3 minutes) of real-world classrooms. The participants then are prompted to identify effective teaching strategies and specific behavioral examples of those strategies by typing in open text boxes. Participant responses are then coded by trained coders for accuracy. Previous research showed a moderate correlation between performance on the VAIL and the observed quality of teaching interactions in a student teaching placement (Wiens, 2014). Additional validity support for using the noticing

framework as an assessment in teacher education is based on evidence that preservice teachers can be trained to become better at noticing (Sherin & van Es, 2005; Star & Strickland, 2008; Stürmer, Seidel, & Schäfer, 2013).

Basing Video Analysis on Understanding Student-Teacher Interactions

Building on the concepts of noticing and expertise, effective assessment of video analysis must be based on a clear vision of effective teaching. The VAIL is also supported by theory and research on teacher-student interactions. These interactions are proximal processes that take place regularly, over an extended time, and serve as an important part of children's development (Bronfenbrenner, 1993). Developed through extensive classroom observations, the Teaching Through Interactions Framework (TTIF) organizes interactions into three domains: Emotional Supports, Classroom Organization, and Instructional Supports (Hamre et al., 2013).

The TTIF provides a framework for understanding teacher-student interactions. It is most often measured using the Classroom Assessment Scoring System (CLASS: Pianta, La Paro, & Hamre, 2008; Pianta & Hamre, 2009) in a standardized and reliable way (Cadima, Leal, & Burchinal, 2010; Graue, Rauscher, & Schefinski, 2009). All three domains of the TTIF have been linked to positive academic outcomes including vocabulary growth (Cadima et al., 2010), phonological awareness (Curby, Rimm-Kaufman, & Ponitz, 2009), reading (Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008), and grades (Reyes, Brackett, Rivers, White, & Salovey, 2012). CLASS has been used as an assessment in teacher education (see Wiens, Hessberg, LoCasale-Crouch, & DeCoster, 2013), but as an assessment of teaching interactions—such as those observed during the student teaching experience—it is most valid late in teacher education programs because that is when preservice teachers have their most authentic teaching experiences.

Originally developed for a large-scale study of pre-kindergarten teachers (Hamre et al., 2012), the VAIL has been implemented as an assessment in teacher education evaluation due to the fact that it can be implemented at multiple points in a teacher education program including as a pretest before preservice teachers even begin their training (Wiens et al., 2013). The VAIL is based on the TTIF framework (Jamil et al., 2015) as it uses videos of in-service teachers selected because they demonstrate the different domains of the TTIF. The VAIL uses the CLASS framework to understand teaching interactions and the VAIL videos match to CLASS domains as shown in Table 1. This study examines longitudinal data collected over five years of administering the VAIL in a teacher education program.

Student Motivation in Teacher Education Assessments

The implementation of the VAIL in teacher education is unique, as it was administered to preservice teachers in their introduction to education course and every subsequent year they were enrolled in the teacher education program. Therefore, not only were we able to assess student scores on the VAIL itself but also student motivation over time. It is important to consider preservice teacher motivation on the VAIL assessment over time, as motivation has been shown to impact academic performance (Dev, 1997) and over-surveying preservice teachers may lead to reduced responsiveness (Porter, Whitcomb, & Weitzer, 2004).

Most teacher education programs use a variety of assessments to evaluate their programs. These assessments can be considered either high stakes for the preservice teachers or low stakes. High-stakes assessments have negative consequences for the individual if they do not pass. For example, many programs use passing rates on licensure exams as one assessment in their evaluations. These are high stakes because if the preservice teacher does not pass then he/she cannot become a licensed teacher. Low-stakes assessments have no potentially negative consequences to the preservice teacher. The VAIL is an example of a low-stakes assessment. The preservice teachers were required to complete the VAIL; however, the results of the VAIL were not reported to the preservice teachers and their performance had no impact on their movement through the teacher education program or ability to become a licensed teacher.

After five continuous years of data collection, this paper examines what lessons can be learned about the assessment and its usefulness for teacher education evaluation. This data-gathering effort is unique in the teacher education field and examining the data provided may inform other teacher education programs in designing their own assessment frameworks.

Table 1
Alignment of CLASS and VAIL domains and dimensions (adapted Piante and Hamre, 2009)

Domains	Pre-K Dimensions	Indicators
Emotional Supports	**Positive climate**	Relationships, Affect, Respect, Communication
	Negative climate	Negative Affect, Punitive Control, Disrespect
	Teacher Sensitivity	Awareness, Responsiveness, Action to Address Problems, Comfort
	Regard for Student Perspectives*	Flexibility, Autonomy, Peer Interactions, Student Expression
Classroom Organization	**Behavior Management**	Clear Expectations, Proactiveness, Redirection
	Productivity	Maximizing Learning Time, Efficient Routines and Transitions
	Instructional Learning Formats*	Learning Targets, Variety of Modalities, Active Facilitation, Student Engagement
Instructional Supports	**Concept development**	Analysis/Reasoning, Creativity, Integration
	Quality of feedback[*]	Feedback Loops, Scaffolding, Building on Responses, Encouragement
	Language modeling	Conversation, Open-endedness, Repetition/Extension, Advanced Language

[*]Diminsions included in the VAIL instrument

In the arena of low-stakes assessments, where there is no external motivation, it may fall on individuals' intrinsic motivation for them to successfully complete the task. "Intrinsic motivation is motivation that is animated by personal enjoyment, interest, or pleasure" (Lai, 2011, p.4). In academics, intrinsic motivation has been linked to task persistence and the amount of time a student will spend on a task (Brophy, 1983). However, a task where the student has little or no interest will generate less intrinsic motivation (Deci & Ryan, 1985; Woolfolk, 1990). Research indicates that students with high levels of intrinsic motivation function more effectively in school (Dev, 1997). Given the connection between intrinsic motivation and academic success it is important to examine the relationship between academic success (in this study grade point average) and both success and effort on the VAIL.

The VAIL, built upon an empirical and theoretical framework that includes teacher noticing and teacher-student interactions, uses video analysis as a means of assessing preservice teachers' knowledge of teaching interactions.

Study Purpose

Teacher education programs seek to find innovative ways to administer assessments that contribute to effective evaluation. In order to address this need, one teacher education program administered the VAIL (Jamil et al., 2015) which conforms to Worrell and colleagues' (2014) call for evaluation "informed by well-established scientific methods" (p.2). In this study we examine the following research questions:

1. Does the ability of preservice teachers to identify effective teaching interactions change over the course of a teacher education program?

2. When taking the VAIL multiple times, do preservice teachers continue to demonstrate equal effort?

3. Are there characteristics that predict either final VAIL scores or final effort on the VAIL?

The five-year experience of the teacher education program can inform the discussion of program evaluation.

Methods

Procedures

Preservice teachers were required to complete the VAIL every year they were in the teacher education program. The VAIL was administered through a website. The first opportunity participants had to take the VAIL was during the first two weeks of their introduction to education course, prior to being enrolled in the education program. Once enrolled in the teacher education program, preservice teachers were required to participate in a data pool (Wiens et al., 2013) where they needed to earn research credits every spring semester. The VAIL was a requirement of the data pool and the preservice teachers could take the VAIL online any time during the spring semester. The online interface for the VAIL is shown in Figure 1. The data pool and the administration of the VAIL were both done by a program-funded doctoral graduate assistant.

Given the connection between intrinsic motivation and academic success, it is important to examine the relationship between academic success (in this study grade point average) and both success and effort on the VAIL.

Every summer the teacher education program paid four doctoral students $1000 each ($4000 total each summer) to code the VAIL responses. The coding team attended a three-hour training session and were required to pass a reliability test with 80% agreement with a master code list prior to beginning coding. Once coding began, the coding team would have weekly drift-check meetings to ensure that coding was reliable. Any time a coder fell under 80% agreement with the master code list that coder would stop coding, retrain, pass a new reliability test, and then resume coding.

Figure 1
VAIL Online Interface

Context and Participants

Data for this study come from a highly selective public university in a mid-Atlantic state. The university has two teacher education programs that lead to teacher licensure: a five-year bachelor's plus master's degree (n=226) and a two-year postgraduate degree (n=48). There are four different programs: early childhood (n=3), elementary (n=114), secondary (n=113), and special education (n=44). Of the participants, 71% were female, 13% male, and 3% unspecified.

Data for this study included all preservice teachers with multiple VAIL scores. For preservice teachers with more than two VAIL scores we used only the first score and the last score. For some bachelor's students the scores may be spread over multiple years. However,

for the post-graduate students the two scores were always in consecutive years, as it is a two-year program. The total number of preservice teachers with multiple years of VAIL scores were 281.

Measures

Video Assessment of Interactions in Learning (VAIL). The VAIL (Jamil et al., 2015) consists of three videos of pre-school language arts classrooms. These videos are followed by prompts instructing participants to identify teaching strategies and specific examples of those strategies from the video. After watching the video, participants had the opportunity to provide five effective teaching strategies they identified from the video in an open-ended format. Examples of effective teaching strategies included in the VAIL would be scaffolding, eliciting student ideas, and variety of instructional modalities.

For each strategy the participant had the opportunity to provide a specific example of the strategy taken from the video. The assessment defines an example as, "a teaching method used to meet a specific goal" (VAIL, 2010). In other words, examples constituted specific actions observed in the video. For example, if a participant noted scaffolding as a strategy a matching example might consist of the teacher helping the student sound out the word the student was struggling to read.

Responses supplied by participants were open ended and were coded for accuracy against a master code list created by master coders. Any differences between coders and the master code list were reconciled based on standards identified in the CLASS (VAIL 2010). The VAIL was designed so that CLASS-specific terminology was not necessary to perform well on the assessment. Participants could use any synonymous terms that identified the teaching strategies indicated in the VAIL manual. The VAIL uses a standardized rating description as outlined in the VAIL Coding Manual (2010) to guide all coding decisions.

To analyze the VAIL data, sum scores were calculated. Previous analysis of VAIL data with in-service teachers presented evidence to support using a one-factor model for compositing VAIL scores using the strategy, example, match and breadth scores (Jamil et al., 2015). The completion variable is analyzed separately because it does not conceptually measure a participant's ability to detect effective teaching interactions; instead, it measures participants' persistence in completing the assessment.

When a CLASS-matched strategy was identified by the participant, a breadth score was also assigned. Each assigned breadth score corresponded to a specific CLASS indicator. The number of unique indicators supplied by participants was then summed to create a breadth score for the entire set of responses for that video. Two of the videos had four possible strategy categories while the third video contained five possible strategy categories. Additionally, if both the strategy and example supplied were correct, the response was coded based on whether the example was an accurate example of the strategy identified.

The completion score measured how many responses the participants wrote for each video. Participants were coded for each attempt at identifying a strategy and example even if the strategy and example were not correctly identified. Each participant was required to provide at least one strategy and example to continue in the assessment. While there was the opportunity to identify five strategies and examples, only one response was required to continue with the assessment. Any strategy-example pairs that were left blank were coded as a zero.

Jamil and colleagues (2015) suggest an analysis strategy that standardizes values within the different videos and then composites the videos into a single score. However, it may be easier to understand the results of the VAIL, particularly when examining longitudinal change, using a sum score. Additionally, using a sum score also facilitates comparison of participant scores across contexts and administrations of the VAIL by providing a fixed number for the final score. The drawback of this approach is that the videos do not all have the same total possible points and therefore one video might have a slightly smaller weight in the overall score than the other videos. The total possible points for the Regard for Student Perspectives video is 19, Instructional Learning Formats is 19 as well, and the Quality of Feedback video

As every teacher education program is should be guided by best practice as well as internal and external stakeholders, finding innovative ways to administer assessments that contribute to the effective evaluation of programs is critical.

total is 20. The differences in possible points comes from the breadth score which has a maximum of four strategies in Regard for Student Perspectives and Instructional Learning Formats, while there are five total strategies in Quality of Feedback. While a sum score makes the Quality of Feedback video slightly more important, the benefits of a sum score outweigh these disadvantages.

Grade Point Average (GPA). The GPA data used in this study was taken from the end of program, cumulative GPA. GPA at this institution is on a four-point scale. The GPA data was taken from administrative records provided by the Teacher Education Office. GPA scores ranged from 2.61 to 4.00. The mean GPA was 3.57 with a standard deviation of .28.

Analysis

For our analysis, we used the first time they took the VAIL (first-test) and the last time they took the VAIL (last-test). The completion score was used as a test of effort. We examined both the VAIL totals and examined the three individual videos. We began with descriptive analysis and correlation estimates to better understand the data. Next, we computed paired sample t-tests to examine differences in variables—particularly focused on examining the differences between first-tests and last-tests. Finally, we computed multiple regression analysis to determine the relationship between effort, the amount of times participants took the VAIL, GPA, and teaching area and VAIL scores and effort.

Results

We began with an examination of the data. Mean scores for GPA and VAIL times taken are in Table 2 while first-test and last-test VAIL and Completion scores are presented in Table 3. Correlations, illustrated in Table 4, indicate that the first- and last- VAIL and first- and last- Completion scores are all significantly correlated with each other. Data analysis did not show a significant difference between first-test and last-test VAIL scores in this sample. We did find that preservice teachers scored higher on the first video than the last video (difference=.303, p=.09); however, this was only significant at the less stringent .1 value. Additionally, t-test analysis found that participants provided fewer responses to the third video last-test (Mean difference=.347, p=.001). In total, participants had lower Completion scores in the last-test than in the first-test (Mean difference= .518, p=.019).

Table 2
Mean Values for Variables

	Mean	SD
VAIL Times Taken	2.04	.20
GPA	3.57	.28

Table 3
Regression Table with Standardized Betas

	Mean (SD)			
	First-test	Last-test	First-test Completion	Last-test Completion
VAIL total	15.99	15.49	12.12	11.60
	(6.69)	(6.96)	(2.95)	(3.12)
Video 1	4.51	4.81	4.10	4.05
	(2.96)	(2.97)	(1.11)	(1.11)
Video 2	5.13	5.03	4.09	5.03
	(2.78)	(3.17)	(1.14)	(3.17)
Video 3	6.21	5.56	3.95	3.60
	(3.39)	(3.64)	(1.25)	(1.57)

Completion scores are all significantly correlated with each other. Data analysis did not show a significant difference between first-test and last-test VAIL scores in this sample.

Our regression analysis is shown in Table 5. When entering the times the VAIL was taken (Times Taken), GPA, and Teaching Area as predictors, we found the overall model to be significant for both last-test VAIL effort (Final R = .218, p=.02) and last-test VAIL scores (Final R = .237, p=.05). Within the regression model predicting final VAIL effort, Times Taken (Standardized β= -.168, p=.01) and Teaching Area (Early Childhood compared to Elementary: Standardized β= -.156, p=.02) were both significantly associated with the VAIL. Within the regression model predicting last-test VAIL score, GPA was the only individual variable that was significant (Standardized β= .172, p=.01).

Table 4
Bi-variate Correlations

	1	2	3	4	5	6
1. First-test VAIL	1	.353***	.501***	.148**	-.083	.076
2. Last-test VAIL		1	.147*	.589***	-.012	.178**
3. First-Completion			1	.270***	-.122*	.010
4. Last-Completion				1	-.137*	.012
5. Times Taken					1	-.064
6. GPA						1

*p < .05
**p < .01
***p < .001

Table 5
Bi-variate Correlations

	VAIL Effort		VAIL Total	
Predictors	ß	Std. Error	ß	Std. Error
Times Taken	-.168*	.865	-.006	2.020
GPA	-.019	.679	.172*	1.586
Teaching Area[a]				
Early Childhood	-.156*	2.074	-.104	4.844
Secondary	-.002	.422	-.071	.985
Special Education	-.067	.565	-.078	.985
Final R	.237*	2.901	.218*	6.775
Final ΔR^2	.056*	2.901	.047*	6.775

[a]Elementary is the comparison group.
*p<.05

Discussion

Many teacher education experts have called for improved teacher preparation instruments that can contribute to efforts to strengthen teacher evaluation (Worrell et al., 2014; Zeichner, 2005). To contribute to one teacher education program's evaluation efforts we examined the use of the Video Assessment of Interactions in Learning (Jamil, et al., 2015) over a five-year period. We found that teacher education students did not demonstrate improved performance on the VAIL from the beginning to the end of their program; however, we did find that participant effort towards the VAIL measure decreased from the beginning to the end of the program. Data indicate that repeatedly expecting teacher education students to take the same assessment may lead to measurement fatigue and lack of effort. Given the cost of implementing the VAIL and the results of overuse, future use of the assessment should be adjusted accordingly.

Worrell and colleagues (2014) call for valid and scientifically based assessments when evaluating teacher preparation programs. An important step in determining the validity of an assessment is understanding the data it provides in practice. The VAIL, which includes watching videos and responding to open-ended prompts, was required of teacher education students every year they were in the preparation program. The VAIL did not show differences

between participants' ability to identify effective teaching interactions at the beginning and end of the program. These might be attributable to the fact that participants demonstrated less effort at the end of their program than at the beginning. The only portion of the VAIL that did show a difference was the first video which also had the most consistent effort of participants at the beginning and end of their program. However, the VAIL does have the benefit of being a standardized measure that can be implemented at various points in the teacher education program (Wiens et al., 2013). It might be advisable to reduce the number of times participants are required to take the VAIL and see if they are more motivated to expend more effort at the end of their program.

Five years of data collection indicate that programs need to carefully consider the burden that their assessments place on participants. Assessing participants too often with the same measure may undermine the validity of the assessment if participants' effort decreases over time.

There appears to be an element of assessment fatigue in our data, as seen in the reduced completion scores in the last-test Completion score compared to the first-test. Assessment fatigue is also supported by the regression analysis that showed a negative relationship between the number of times a participant took the VAIL and the effort he/she was willing to put into the final attempt. Even within the assessment the third and final video had the lowest completion score, and in the last-test the third video also had the lowest completion of any video from any time point. In this teacher education program the VAIL is a low-stakes assessment and it relies on preservice teachers' intrinsic motivation to do well. Since intrinsic motivation is related to personal enjoyment, interest, or pleasure (Lai, 2011), it might be difficult to motivate students to do their best work. While there is little empirical literature related to assessment fatigue in these situations, there is evidence that university students who are expected to complete multiple surveys may be unlikely to participate fully (Porter et. al., 2004) and this may be especially true in longitudinal surveys (Apodaca, Lea, & Edwards, 1998). In this sample, fatigue may be an issue due to low motivation and repeated administrations of the same measure; the more times teacher education students were asked to complete the VAIL the less effort they were willing to put into completing the measure.

The VAIL has been shown to be a valid and reliable measure (Jamil et al., 2015), related to teaching performance with in-service (Hamre et al., 2012) and preservice teachers, and useful in teacher education contexts (Wiens et al., 2013). However, this study provides some important information for determining best practices for use of the VAIL as a teacher education program evaluation tool. The VAIL gives participants the opportunity to provide up to 30 different responses (five strategies plus five examples for each of the three videos). Future implementation of the VAIL should revisit the length or layout of the measure to make it a more valid estimate of preservice teachers' ability to identify effective interactions. Another option is to require teacher education students to only take the VAIL at the beginning and end of the program to determine if participant effort improves on the last-test. This would also have the benefit of requiring less resources from the teacher education program in hiring and training reliable coders. A third option to increase participant effort on the VAIL would be to experiment with making it a higher-stakes assessment. If participants were more motivated to do well on the assessment then they may increase their effort and improve their overall performance.

Conclusion

Systematic research on teacher education is a necessity for the field (Grossman & McDonald, 2008; Worrell et al., 2014; Zeichner, 2005). The development of valid measures (Worrell et al., 2014) that can address the needs of multiple constituents (Feuer et al., 2013) can help to move the field forward and provide robust program evaluations. One teacher education program used the VAIL (Jamil et al., 2015) as a component of its evaluation program. Five years of data collection indicate that programs need to carefully consider the burden that their assessments place on participants. Assessing participants too often with the same measure may undermine the validity of the assessment if participants' effort decreases over time. Continual examination of program assessments is required to ensure that teacher education programs are preparing future generations of quality teachers.

References

Apodaca, R., Lea, S., & Edwards, B. (1998). The effect of longitudinal burden on survey participation. In *American Statistical Association Proceedings of the Survey Research Methods Section* (pp. 906–10).

Bransford, J.D., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bronfenbrenner, U. (1993). Ecological models of human development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children, 2nd ed*. (pp. 37–43). New York, NY: Freeman.

Brophy, J. (1983). Conceptualizing student motivation. *Educational Psychologist, 18*(3), 200–215.

Cadima, J., Leal, T, & Burchinal, M. (2010). The quality of teacher-student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology, 48*, 457–482. doi: 10.1016/j.jsp.2010.09.001

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education, 39*(3), 25–31.

Christ, T., Arya, P., & Chiu, M.M. (2017). Video use in teacher education: An international survey of practices. *Teaching and Teacher Education, 63*, 22–35. doi:10.1016/j.tate.2016.12.005

Council for the Accreditation of Educator Preparation. (2013). *2013 CAEP standards*. Washington, DC: Author.

Curby, T.W., Rimm-Kaufman, S.E., & Ponitz, C.C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*(4), 912–925. doi: 10.1037/a0016647

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

Dev, P.C. (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? *Remedial and Special Education, 18*(1), 12–19.

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.

Feuer, M.J., Floden, R.E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education.

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41–67. doi:10.1016/j.edurev.2015.06.001

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

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Graue, E., Raushcer, E., & Sherfinski, M. (2009). The synergy of class size reduction and classroom quality. *The Elementary School Journal, 110*(2), 178–201. doi: 10.1086/605772.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184–205. doi: 10.3102/0002831207312906

Jamil, F.M., Sabol, T.J., Hamre, B.K., & Pianta, R.C. (2015). Assessing teachers' skills in detecting and identifying effective interactions in the classroom: Theory and measurement. *The Elementary School Journal, 115*(3), 407–432.DOI: https://doi.org/10.1086/680353

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (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. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (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

Lai, E. R. (2011). *Motivation: A literature review*. Boston, MA: Pearson.

Pianta, R.C., La Paro, & Hamre, B.K. (2008). Classroom Assessment Scoring System (CLASS). Baltimore, MD: Paul H. Brookes.

Pianta, R.C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F.J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*, 365–397. doi: 10.3102/0002831207308230

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), pp. 109–119. doi: 10.3102/0013189X09332374

Porter, S. R., Whitcomb, M.E., & Weitzer, W.H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research, 121*, 63–73.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*(3), 700–712. doi: 10.1037/ a0027268

Santagata, R. & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching   effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education, 17*, 491–514. doi: 10.1007/s10857-013-9263-2

Sherin, M.G. & van Es, E.A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education, 13*(3), 475–491.

Star, J.R. & Strickland, S.K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education 11*(2), 107–125. doi: 10.1007/s10857-007-9063-7

Stürmer, K., Seidel, T., & Schäfer, S. (2013). Changes in professional vision in the context of practice. *Gruppendynamik und Organisationsberatung, 44*(3), 339–355. doi: 10.1007/s11612-013-0216-0

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

Video Assessment of Interactions and Learning: VAIL (2010). Coding manual (Unpublished Manual). Charlottesville, VA: Center for the Advanced Study of Teaching and Learning.

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.

Wiens, P.D. (2014). Using a participant pool to gather data in a teacher education program: The course of one school's efforts. *Issues in Teacher Education, 23*(1), 177–206.

Woolfolk, A. E. (1990). *Educational psychology* (4th ed.). Boston: Allyn & Bacon.

Worrell, F., Brabeck, M., Dwyer, C., Geisinger, K., Marx, R., Noell, G., & Pianta R. (2014). *Assessing and evaluating teacher preparation programs*. Washington, DC: American Psychological Association

Zeichner, K. M. (2005). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 737–759). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; Washington, DC, US: American Educational Research Association.