

RESEARCH CONFERENCE 2020

Research Leading the Way For The Next 50 Years

**Student Voices in Teacher Evaluation: A
Multilevel, Latent Factor Investigation of
Teacher Quality**

**POSTER
PRESENTATION**

OCTOBER 1, 2020

**WALDEN
UNIVERSITY**

Author

Kathleen J. Hoff, PhD

Graduate

School of Psychology

Kathleen.j.hoff@gmail.com

LinkedIn: www.linkedin.com/in/jodyhoff

Dissertation Committee

Steven Kriska, Committee Chair

Neal McBride, Committee Member

Melody Moore, University Reviewer

Abstract

Teacher effectiveness is a key driver of student achievement but difficult to measure. The purpose of this study was to investigate the validity of the Tripod Survey, the most widely used precollege survey in the U.S., as a measure of teacher effectiveness. Multilevel confirmatory factor methods were used to evaluate the factor structure of the survey. None of the specifications adequately fit the data, suggesting the Tripod survey did not reflect effective teaching as hypothesized by the survey authors. These results provide educational leaders with more robust information about teacher evaluation surveys and can lead to more informed decision making.

Doctoral Dissertation

Problem

Teachers represent the single most important factor in determining student achievement that can be controlled by a school district

Teaching is complex, nuanced, and perplexingly difficult to measure.

Precollege student surveys are a new and growing form of teacher evaluation.

Little empirical evidence exists about the validity of precollege surveys to measure teacher effectiveness.

Purpose

The purpose of this quantitative study was to explore the viability of a student survey to serve as a valid and reliable indicator of teacher quality.

Significance

Incorporating student perceptions of the learning environment into the evaluation process may enhance discrimination of teacher effectiveness and provide insights for instructional improvement (Goe, Bell, & Little, 2008; Scherer, Nilsen, & Jansen, 2016).

Evaluations provided by student ratings were potentially a useful tool for informing and improving instructional practices of individual teachers.

Adding the student voice to the evaluation process offers valuable and rarely heard insights into what constitutes effective teaching (Chaplin, Gill, Thompkins, & Miller, 2014; English et al., 2015; Martínez, Schweig, & Goldschmidt, 2016; Ravitch, 2013).

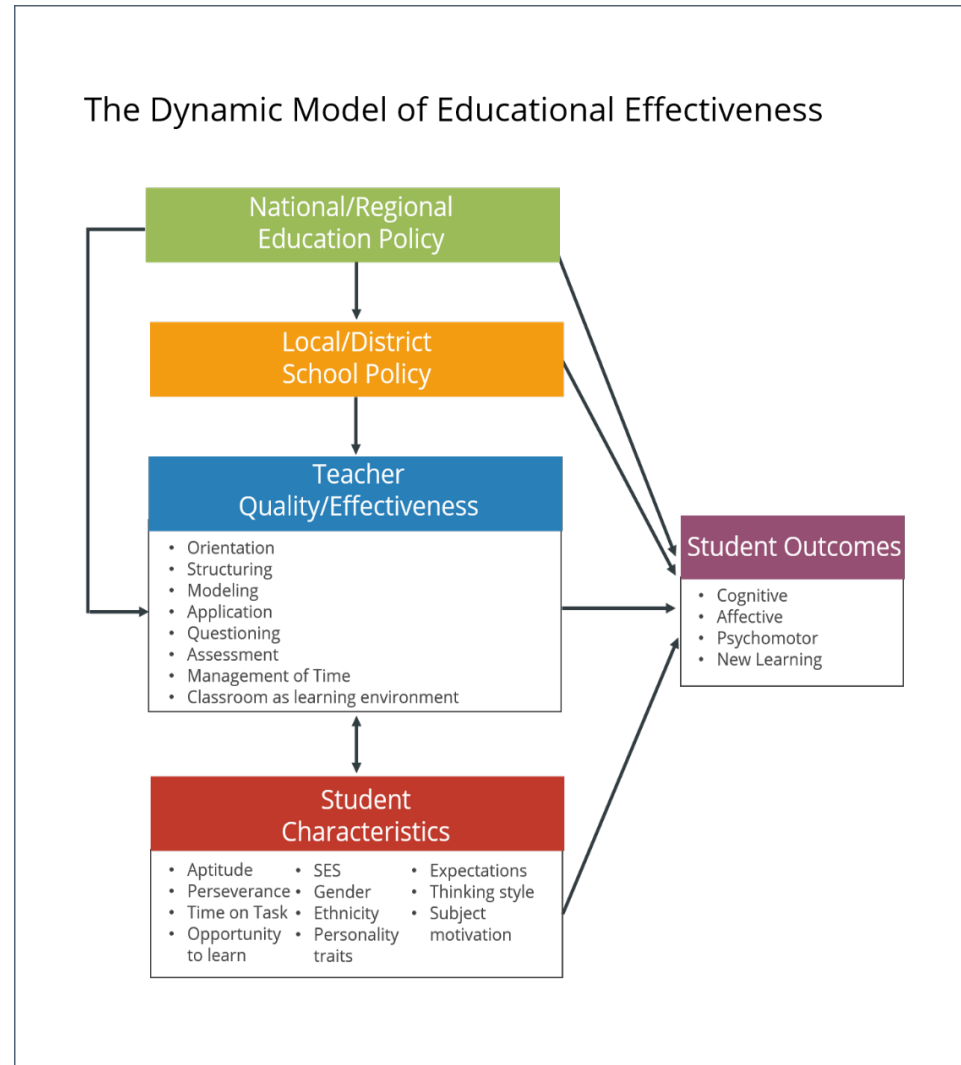
The first hurdle in establishing such a scale is evaluating the psychometric properties of the survey qualities from a multilevel construct validation perspective (Jebb et al., 2019; Scherer et al., 2016; Tay et al., 2014), which this study addressed.

Macro-framework

The **Dynamic Model of Educational Effectiveness (DMEE)** is a comprehensive theoretical framework used to evaluate factors that impact school-based learning (Creemers, 2002; Creemers et al., 2013).

The DMEE incorporates key aspects of the educational environment, including (a) the preeminence of the relationship between teacher and student, (b) multidimensional aspect of instructional practices and characteristics, (c) multilevel analysis indicative of school settings, and (d) the existence of direct and indirect effects of instructional practices on student outcomes (Kyriakides & Creemers, 2009; Reynolds et al., 2014).

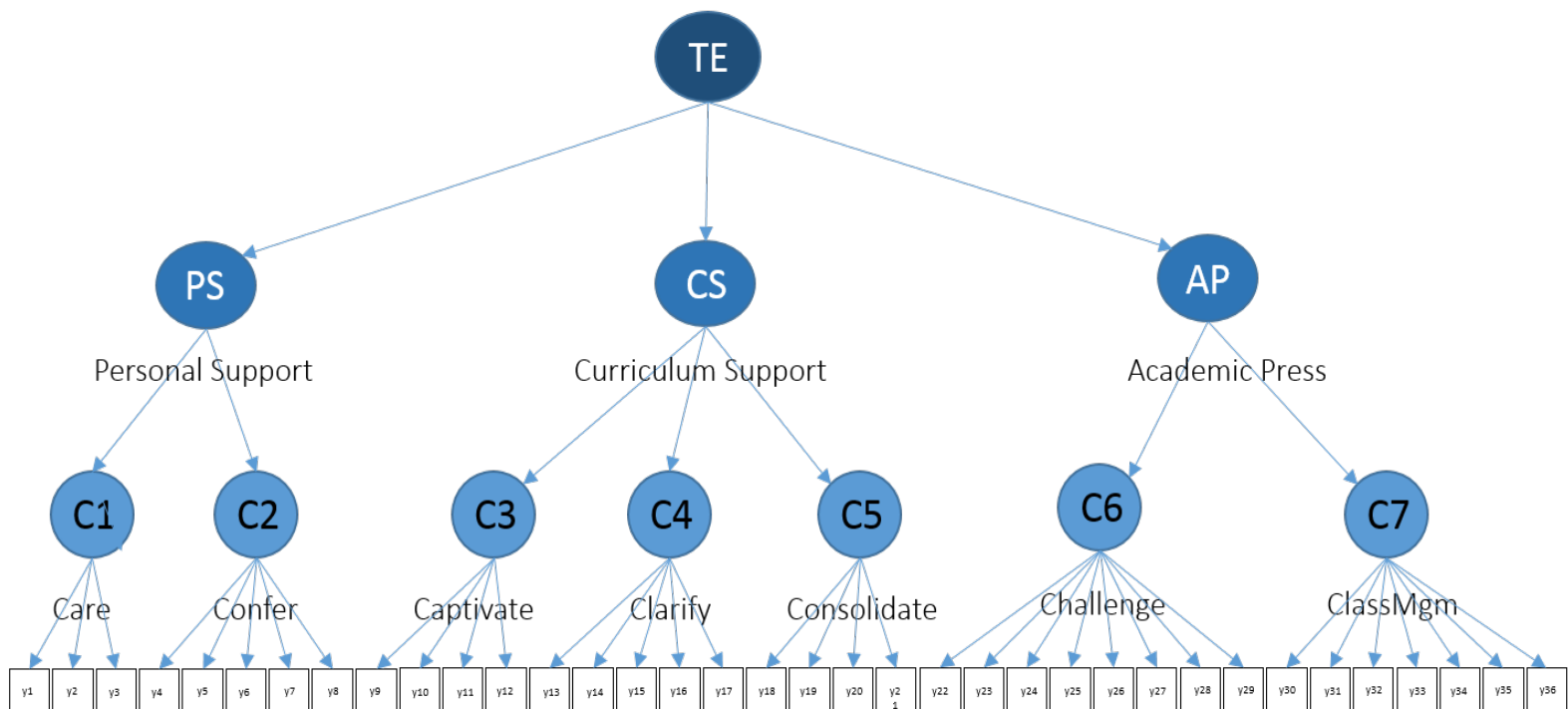
These four aspects highlight the proximity of the theory itself to the realities of school-based learning.



Micro-framework

The Tripod Student Perception Survey (Ferguson, 2012) contains seven unique factors that form the foundation of teacher effectiveness as measure by 36 individual items administered to students. Each of the seven factors are also hypothesized to measure three subscales.

Tripod 7Cs Conceptual Framework Teacher Effectiveness



Relevant Scholarship

More than any other factor that can be controlled by a school district, a teacher's effectiveness in the classroom is the most important in determining student achievement (Chetty, Friedman, & Rockoff, 2014).

Measuring the effectiveness of k-12 teachers is a critical component in improving outcomes for student learners (Harris & Sass, 2011).

Although of last majority of teachers are consistently rated as highly qualified nationwide, corresponding student achievement scores continue to vary widely across individual states and schools (Huber & Skedsmo, 2016).

School districts across the country have undertaken large-scale efforts to rethink and rebuild teacher evaluation systems (Douglas N Harris, Ingle, & Rutledge, 2014).

Evaluating the effectiveness of K-12 teachers is expensive and time-consuming (Rockoff & Speroni, 2010).

Although commonly used at the post-secondary level, student surveys at the primary and secondary levels represent a new, inexpensive, and relatively untested component of teacher evaluation systems (Geiger, Amrein-Beardsley, 2019; .

Analysis by Schweig (2014b), Flenor (2015), Kuhfeld (2016), and Wallace et al. (2016) also had mixed results and found no evidence to support the 7-factor model proposed by the Tripod authors.

Further, Schweig (2014b) found inconsistencies across theoretical models using the Tripod and the importance of justifying the reliability of teacher effectiveness scores.

Research Questions

RQ1: What is the factor structure of the Tripod Student Perception Survey (Tripod) at the student level?

RQ2: Does teacher effectiveness, as measured by the Tripod, represent a multilevel construct?

RQ3: To what extent does the construct of teacher effectiveness, as measured by the Tripod, exhibit a common factor structure across the student and classroom levels (i.e., psychometric isomorphism)?

RQ4: Does the Tripod exhibit a consistent factor structure across measurement periods?

Participants

21,000 9th grade students from six large, urban school district in the U.S., studying language arts, mathematics, and science. These student rated the effectiveness of their classroom teacher as part of the Measures of Effective Teaching study.

Data Source

Data were obtained from the **Measures of Effective Teaching Longitudinal Database**, which is available as a restricted dataset via the Inter-University Consortium for Political and Social Research at the University of Michigan.

Stata and Mplus software were used to process data and complete multilevel factor analysis.

Analysis

A multilevel model evaluation strategy to determine the degree of construct isomorphism present in a series of hypothesized and empirically generated factor models based on the Tripod Student Perception Survey student responses.

Findings

No empirical evidence was found in support of the hypothesized seven-factor structure of the Tripod survey as stated by the survey authors.

No empirical evidence was found at the classroom level in support of alternative factor structures suggested by other researchers, including the higher-order bifactor models.

A two-factor model comprising the individual factors of Captivate and Clarify emerged as a potential measure of teacher effectiveness at both levels.

The reserve coded items, consistent with the findings from Wallace et al. (2016), proved challenging for students to interpret.

The Year 1 results for the two-factor model and the reverse coded items were replicated in the Year 2 data.

Interpretation

The empirical evidence suggests that the 36 items that compose the Tripod instrument do not capture effective teaching as hypothesized.

The Tripod seven-factor model and its variants appeared to not index a shared perspective from students that could be generalized to classroom teachers.

Without supporting validity evidence, the theory of effective instruction, as operationalized by the Tripod, does not appear to be supported in the MET data for Grade 9 students.

A bifactor specification that had previously been reported as a potential alternative (Wallace, 2016).

A two-factor reduced-form of the Tripod appeared to hold some promise of explaining teacher effectiveness through the perspective of student raters.

Limitations

Convenience sampling and the attrition of teachers between Y1 and Y2 hindered the generalizability of results.

Study design did not include a standardized test for the Grade 9 students, which limited the ability to validate findings against an outside criterion, such as a standardized test.

The inconsistent use of language to describe construct validation of multilevel models can be challenging (e.g., the term isomorphism to describe cross-level similarities between constructs).

Recommendations

Work in multilevel construct validation is evolving (Tay et al., 2014). Combining (a) partially saturated methods (Rhy & West, 2009) for determining configural isomorphism and (b) evaluating the equality of loadings (Jak, 2013, 2014) to establish metric isomorphism can provide a more complete picture of psychometric isomorphism.

Given the factor structure at the classroom did not correspond to the same structure at the student level, it would be worthwhile to explore what factor structure actually exists at the level of the classroom.

Investigate the functioning of individual items, including item discrimination and difficulty, to more thoroughly understand student responses.

Carefully consider the use of reverse-coded items as student were confused about these items.

Social Change Implications

Education is arguably, the single largest agent of social change.

Of all the resources a school can directly control, the classroom teacher matters most for student outcomes.

Measuring teacher quality has been and continues to be challenging.

The most widely used pre-college student survey in the U. S. does not appear to measure teacher quality as hypothesized.

Policy makers need information about the validity of evaluation instruments to ensure teachers are appropriately evaluated and ensure every student has access to high quality instruction.

References

- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632. <https://doi:10.1257/aer.104.9.2593>
- Ferguson, R. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, 94(3), 24–28. [https://doi: 10.1177/2F003172171209400306](https://doi:10.1177/2F003172171209400306)
- Geiger, T., & Amrein-Beardsley, A. (2019). Student perception surveys for K-12 teacher evaluation in the United States: A survey of surveys. *Cogent Education*, 6(1). [https://doi: 10.1080/2331186x.2019.1602943](https://doi:10.1080/2331186x.2019.1602943)
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7–8), 798–812. <https://doi:10.1016/j.jpubeco.2010.11.009>
- Jak, S. (2019). Cross-level invariance in multilevel factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(4), 607–622. <https://doi:10.1080/10705511.2018.1534205>
- Kuhfeld, M. (2017). When students grade their teachers: A validity analysis of the Tripod student survey. *Educational Assessment*, 22(4), 253–274. <https://doi:10.1080/10627197.2017.1381555>
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review*, 100(2), 261–266. <https://doi:10.1257/aer.100.2.261>

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, 16(4), 583–601. <https://doi:10.1080/10705510903203466>

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the Tripod Student Perception Survey. *American Educational Research Journal*, 53(6), 1834–1868.
<https://doi:10.1080/02671522.2015.1027724>

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: New Teacher Project. Retrieved from https://tntp.org/assets/documents/TheWidgetEffect_execsummary_2nd_ed.pdf