Scoring with Confidence: Enhancing Preservice Teachers' Ability to Evaluate Instruction

Ashlyn Smith
Texas Tech University
Box 42131-Lubbock, TX 79409-2131
Ashyn.r.smith@ttu.edu

Krysti Kelley
Texas Tech University
Box 42131-Lubbock, TX 79409-2131
krystkel@ttu.edu

William Doss
Texas Tech University
Box 42131-Lubbock, TX 79409-2131
Will.Doss@ttu.edu

Introduction

Preservice teachers undergo significant development through coursework, student teaching experiences, and mentoring from cooperating teachers and university supervisors (McKim & Velez, 2017). Throughout this process, teacher reflection and self-efficacy have been identified as important factors related to teacher persistence, retention, and resiliency (Yost, 2006). Teacher self-efficacy (Blackburn & Robinson, 2008) and job satisfaction (Clemons & Lindner, 2019) among early-career agriculture teachers are critical factors that influence retention and career longevity in the profession. Self-evaluation enables educators to engage in self-observation, self-analysis, and self-reflection, helping them explore their experiences and improve their professional practice (Asad Juma, 2024). However, there is currently no standardized evaluation training, particularly self-evaluations, within teacher preparation programs. The purpose of this study was to evaluate the effectiveness of an intervention designed to train preservice teachers to provide accurate feedback to develop self-reflective practices. It was hypothesized that the accuracy of feedback provided by preservice teachers would significantly improve after instruction. The null hypothesis stated that there is no significant difference in feedback accuracy provided by preservice teachers before and after instruction.

Conceptual Framework

This study integrates components from two established theoretical models. Lipnevich et al.'s (2016) Student-Feedback Interaction Model outlines eight characteristics of effective feedback. Two of these feedback characteristics were utilized in the design of this study: accuracy, referring to the correctness of the feedback provided, and congruency with expectations, defined as the alignment between the feedback received and what the learner anticipated (Lipnevich et al., 2016). These characteristics are essential for understanding how preservice teachers perceive and process information during evaluation. Bandura's (1977) Self-Efficacy Theory describes four sources of experience and highlights mastery experiences as the most influential source of self-efficacy, based on personal experiences of success and failure (Bandura, 1977). This study aimed to provide preservice teachers with practice in giving accurate feedback, serving as mastery opportunities that strengthen their self-efficacy in self-evaluation. By combining these frameworks, a gap in preservice teacher training is addressed: the development of accurate self-evaluation skills through structured experiences.

Methodology

This study used a one-group pretest-posttest design (Fraenkel, et al., 2023) to examine how an instructional intervention affected preservice teachers' feedback accuracy. Eighteen preservice agriculture education students enrolled in an introductory teaching course at Texas Tech University were purposively selected for this study. Threats to internal validity, such as history, maturation, testing, and instrumentation, were limitations of this design. Additionally, purposive sampling limited generalizability beyond the study population, and the absence of a control group limited causal interpretations (Fraenkel et al., 2023). Preservice teachers first watched a 45-minute recorded lesson and used the provided observation instrument to evaluate the educator's performance (pretest). Following this, a 2-hour calibration session introduced participants to the instrument's domains, key indicators, and expert scoring criteria (treatment). After receiving feedback and discussing discrepancies between their ratings and the expert score,

participants viewed a second lesson video and scored the teacher using the same instrument (posttest). The observation instrument used was the Texas Tech University Student Teacher Evaluation Tool, which measures teaching performance across four domains: (1) Designing Instruction and Assessment; (2) Creating a Positive, Productive Classroom Environment; (3) Implementing Effective, Responsive Instruction; and (4) Fulfilling Professional Responsibilities. The instrument included 21 Likert-scale items (0 = *not observed* to 4 = *clearly outstanding*) and demographic questions. Face validity was established by a panel of agricultural education experts. Kelley and Rayfield (2025) reported acceptable reliability for Domain 1 (α = .73), Domain 2 (α =.85), Domain 3 (α = .72), and Domain 4 (α = .76). Two coders independently scored teaching videos (pretest and posttest) to create the expert scores against which preservice teacher scores were compared. Initial Krippendorff's α ranged from .66 to .83; after coder calibration, final coefficients met or exceeded α = .80 (Domain 1 = .91, Domain 2 = .83, Domain 3 = .84, Domain 4 = .81). Data were analyzed using IBM SPSS Statistics. Accuracy difference scores were calculated by taking the absolute value of the difference between participant and expert ratings. Normality of scores was examined using the Shapiro–Wilk test (Field, 2024). Because several variables violated normality assumptions, and the sample size was small, nonparametric Wilcoxon signed-rank tests were used to compare pre- and post-training accuracy within each domain. Effect sizes (r) were calculated, and a Bonferroni adjustment (α = .0125) was applied.

Results/findings

From pretest to posttest, mean difference scores between participant and expert ratings decreased slightly in Domain 1 (M = 0.39, SD = 0.34 to M = 0.32, SD = 0.34) and markedly in Domain 3 (M = 0.56, SD = 0.28 to M = 0.33, SD = 0.24), while Domain 2 showed a small increase (M = 0.33, SD = 0.32 to M = 0.39, SD = 0.26) and Domain 4 remained unchanged (M = 1.91, SD = 0.82 to M = 1.91, SD = 0.82). A statistically significant improvement in scoring accuracy was found for Domain 2 ($Z = -3.75$, $p < .001$, $r = .88$). Changes in Domains 1 (p = .32), 3 (p = .02), and 4 (p = 1.00) were not significant after Bonferroni correction. Lower difference scores indicate greater accuracy between preservice teacher and expert ratings.

Conclusions, Implications, and Recommendations

Training produced a significant improvement in scoring accuracy for Domain 2, suggesting that it was more effective for some evaluation areas than others. This suggests that providing opportunities for accurate feedback, guided by Lipnevich et al.'s (2016) Student-Feedback Interaction Model, effectively supports the development of some self-evaluation skills. Accuracy improvements were not consistent across all domains, suggesting that a single training session may not fully prepare preservice teachers to evaluate all dimensions of teaching performance. This highlights the need for ongoing practice to support preservice teachers' self-evaluation and reflective growth (Asad Juma, 2024). Teacher preparation programs should incorporate evaluation training into their curriculum to promote effective self-evaluation (Asad Juma, 2024). We recommend calibration be more targeted, and future training should allocate more time or examples to promote further opportunities for mastery experiences (Bandura, 1977). It would also be beneficial to offer periodic recalibration sessions to maintain scoring consistency over time. Future research should replicate this study with a larger sample to confirm found effects and assess generalizability. Calibration training effectiveness should also be tested on cooperating teachers to determine if it improves the reliability and consistency of their evaluations of their student teachers.

References

Asad Juma, A. (2024). Self-reflection in teaching: A comprehensive guide to empowering educators and enhancing student learning. *International Journal of Science and Research Archive, 12*(1), 2835–2844. https://doi.org/10.30574/ijsra.2024.12.1.1113

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Blackburn, J. J., & Robinson, J. S. (2008). Assessing teacher self-efficacy and job satisfaction of early career agriculture teachers in Kentucky. *Journal of Agricultural Education*, *49*(3), 1–11. https://doi.org/10.5032/jae.2008.03001

Clemons, C. A., & Lindner, J. R. (2019). Teacher longevity and career satisfaction in the secondary agricultural education classroom. *Journal of Agricultural Education, 60*(1), 186–201. https://doi.org/10.5032/jae.2019.01186

Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2023). *How to design and evaluate research in education* (11th ed.). McGraw Hill.

Kelley, K., Rayfield, J. (2025, May 19-22). *Feedback and Performance: Investigating the Link Between Lesson Evaluations and Final Teaching Assessments*. [Poster presentation]. AAAE National Conference. Lubbock, TX, United States. https://shorturl.at/0GWIZ

Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 169–185). Routledge. https://www.anastasiyalipnevich.com/wp-content/uploads/2020/01/Lipnevich_EECE_2016-Model.pdf

McKim, A. J., & Velez, J. J. (2017). Developing self-efficacy: Exploring preservice coursework, teaching, and professional development experiences. *Journal of Agricultural Education*, *58*(1), 172–285. https://doi.org/10.5032/jae.2017.01172

Yost, D. S. (2006). Reflection and self-efficacy: Enhancing the retention of qualified teachers from a teacher education perspective. *Teacher Education Quarterly*, *33*(4), 59–76. http://www.jstor.org/stable/23478871