



OPEN ACCESS

EDITED BY

Beverly J. Irby,
Texas A and M University, United States

REVIEWED BY

Anikó Fehérvári,
Eötvös Loránd University, Hungary
Antonio Luque,
University of Almeria, Spain
Waldemar Jędrzejczyk,
Częstochowa University of Technology,
Poland

*CORRESPONDENCE

Susan K. Brondyk
✉ brondyk@hope.edu

RECEIVED 15 July 2023

ACCEPTED 26 December 2023

PUBLISHED 23 January 2024

CITATION

Brondyk SK (2024) Engaging in continuous improvement: implications for educator preparation programs and their mentors. *Front. Educ.* 8:1259324. doi: 10.3389/educ.2023.1259324

COPYRIGHT

© 2024 Brondyk. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Engaging in continuous improvement: implications for educator preparation programs and their mentors

Susan K. Brondyk*

Education Department, Hope College, Holland, MI, United States

This is a mixed method study in which I examined how one Educator Preparation Program designed and used an assessment tool to promote growth in their teacher candidates. The tool became central to the identity of the program and foundational to the work of the mentors. Data includes the statistical analysis of seven semesters of assessment data, along with qualitative interviews, recordings of mentor training sessions, three-way mentor conversations, and written materials. Research questions include: How do tools contribute to the identity of a program question? What value does using an internally-designed evaluation tool add to an educator preparation program? The findings indicate that the practice of the mentors became more developmentally-oriented over time. This suggests that engaging in the process of creating, adapting, and using an internally-designed instruments contributes to a shared identity that reflects the values of the program and provides direction to all.

KEYWORDS

mentoring, assessment, program identity, coherence, continuous improvement

Introduction

Mentoring involves a more experienced or knowledgeable person helping another develop skills, abilities, knowledge, and/or thinking. It occurs in many different contexts (e.g., schools, community, businesses, hospitals), but no matter where it takes place, mentors are expected to adhere to the values and goals of the organization or institution which they represent while also employing their professional judgment as it benefits their mentees (Burden et al., 2018). Direction for their work comes in the form of training sessions, conceptual frameworks, practices and tools. Even with these guiding structures in place, variation naturally occurs as mentors need to be free to be responsive to the specific needs of their mentees.

In teacher education, this symbiotic relationship between adherence and variation is critical to successful mentoring as it provides direction to mentors while simultaneously allowing them to use their professional judgment to vary their practice from one student teacher to another (Brondyk, 2020). It consists of program structures that guide mentors' work, like beliefs about teaching and learning, mentoring knowledge/skills and program tools. Mentors are expected to understand, follow and use these as they help emerging teachers learn to teach.

Brondyk (2020) found that mentors do not always act in ways that align with the programs for which they work. There are many possible reasons for this ranging from lack of preparation (Ulvik and Sunde, 2013) to differing beliefs about mentoring, teaching and learning (Lejonberg et al., 2015). In some instances, programs have underdeveloped visions of mentoring, but in

other cases they fail to articulate and/or communicate their mission, values and goal to mentors which limits their ability to adhere to the program. Without structures and preparation, mentors often do as they see fit, which may or may not align with the intentions of the program. In this study, I examined the role that one assessment tool played in shaping an Educator Preparation Program's (EPP) identity and in particular, the work of its mentors.

Program description

The context for this study is a small liberal arts college in the Midwest. The Education Department has 350 students and graduates approximately 75 students per year and offers certification tracks in elementary (with 13 majors and minors), secondary (with 40 majors and minors), and K-12 programs in Special Education, Early Childhood, English as a Second Language, Visual Arts, Music Education and Physical Education/Health. The program has been nationally accredited since 1960, first through the National Council for the Accreditation of Teacher Education (NCATE), then earning accreditation in 2012 through the Teacher Education Accreditation Council (TEAC), and most recently through the Council for the Accreditation of Educator Preparation (CAEP) Accreditation for 7 years.

Program graduates are sought after both locally and nationally with placement rates for graduates with recent figures standing at a 100% placement rate within 3 months of graduation in their content area. Surveys done on the performance of graduates indicate that they are performing exceptionally well in the field, with more than 95% of graduates rated as "Highly Effective" or "Effective" on the statewide rating system during their first 3 years in the field.

Student teaching occurs during the final semester of the program and lasts 16 weeks. Depending on the major, candidates are placed in one classroom except for those with K-12 majors who split the time in multiple placements. Student teachers are mentored by both a cooperating (P-12) teacher and college supervisor. Cooperating teachers not only open their classroom and teaching to the student teacher, but also fill various roles, including mentor, model, evaluator, trusted listener and resource person. Throughout the semester, cooperating teachers help move their teacher candidate forward by making their thinking explicit and coordinating with the college supervisor as they co-mentor the teacher candidate. College supervisors play a critical role as liaisons between the program, student teacher, and cooperating teacher. They are charged with oversight of the student teaching experience, officially observing candidates a minimum of eight times, working with candidates individually (as needed), leading weekly seminars, and ultimately evaluating their performance.

Impetus for change

The program's 2012 TEAC accreditation process resulted in an Area for Improvement assigned to their evaluation form, which was used to evaluate candidates both at midterm and end of the semester. Reviewers felt that the format of the tool encouraged mentors to rate student teachers consistently in the highest range, thus resulting in over-inflated ratings. These findings caused the program to critically

examine the tool and make changes that would result in more accurate and reasonable ratings overall. Ultimately, a committee, called the Performance Evaluation Committee (PEC), was formed within the Education Department and tasked with redesigning the student teaching evaluation form. This design-process initiated conversations about learning to teach and sparked ideas that eventually had ramifications beyond the form itself and was eventually a major instigator in overhauling the entire student teaching program.

Redesign process

Articulating beliefs about learning to teach

Work on the evaluation form caused members of the program to initially step back and consider their normative view of learning to teach. Two fundamental constructs quickly became apparent. The first is the belief that learning -- whether to teach or to mentor -- is a life-long, developmental process (Feiman-Nemser, 2001). Learners come to the process with differing experiences, knowledge and skills. Therefore, helping them to develop their practice requires meeting them "where they are" and creating experiences in which they can learn and practice new skills with assistance, as they work their way toward independence (Brondyk and Cook, 2020). Another premise that emerged is that learning is a collaborative endeavor, as program members believe that people learn from one another. As such, conversations are key components to learning to teach, especially more intentional conversations where mentors not only make suggestions, but also ask probing questions, inquire about teaching, and make their thinking explicit in order to create habits of mind in their student teachers. These conversations occur formally during 3-way conferences with mentors working together to make moves that would be growth-producing for the student teacher (Brondyk and Cook, 2020).

Creating a new tool

The PEC first set out to make minor revisions to the existing evaluation form, but quickly realized that deeper changes were necessary in order to truly address the issues raised by TEAC. The work of the PEC continued over a 3-year period during which they analyzed the evaluation form's format, scoring rubric, organization and wording. One insight was that the Likert scale format encouraged mentors to simply go down the list and check the "Excellent" box for each indicator without much regard, if any, to the accompanying scoring rubric (Brondyk and Cook, 2020). It became evident that the entire tool would need to be overhauled in ways that would address the over-inflation issue and to also better reflect the program's developmental and collaborative stances.

The committee began by examining existing educator evaluation systems and ultimately decided to loosely frame the new tool around Danielson (2011) *A Framework for Teaching* as it was being used by many districts in the area. In terms of format, they then melded elements of Danielson's framework (domains indicated by color-coding) with the department's existing Professional Abilities (Ethical Educator, Skilled Communicator, Engaged Professional, Curriculum Developer, Effective Instructor, Decision Maker, and Reflective Practitioner) which had been used since 1994 and were integrated into all elements of the program including promotional materials, course syllabi, and, of course, the evaluation form. The committee felt that the

Abilities embodied qualities that the program wanted to instill in their completers (e.g., to be an effective instructor) and were therefore a valuable aspect of the program.

One of the major revisions to the tool related to the proficiency levels. The PEC initially decided on a 3-point system with the categories of Meets Expectations, Developing Expectations and Does Not Meet Expectations, using language that would reflect the department's developmental mindset. Conversations about how to handle teacher candidates who performed above and beyond the Meets Expectations level led to consideration of a fourth level, Exceeds Expectation. They were cautious about this addition, being acutely aware of the AFI from TEAC and wanting to avoid a situation where the mentors felt pressured to inflate ratings when evaluating their student teachers. Ultimately, the committee decided to keep the category because they wanted a way for mentors to be able to acknowledge when student teachers excelled. Instead, they chose to leave the column blank – with no descriptors – recognizing that there might be multiple ways in which a candidate could exceed in any given category. This then required mentors to write in an explanation of how the student teacher had gone beyond meeting expectations, which was considered the target and thus “A” work.

Another part of the process related to the name of the instrument. Discussions about the tool led the group to conclude that they wanted it to be used primarily for assessment rather than merely to evaluate student teachers, hence the name switch from an evaluation form to an assessment tool. The latter more accurately reflected the developmental nature of the student teaching experience as the tool allowed mentors to track the growth of candidates, indicating strengths and growth areas on which they could focus their work.

The result of this process was the Student Teaching Assessment Tool (STAT), a multi-faceted rubric that allowed for assessment of a range of skills and dispositions during student teaching. STAT was designed to be used formatively by the cooperating teacher, college supervisor and student teacher throughout student teaching, allowing each individual to identify the student teacher's level of performance, which then encouraged and facilitated substantive conversations about the candidate's practice and areas for growth. A hardcopy Working Version of the tool was used by both mentors throughout student teaching to track growth. Triads then met regularly to discuss the candidate's development, using STAT as the basis for conversations about effective teaching. The idea was that the mentors would coordinate their efforts to help the student teacher develop in a focal area, rather than working at cross-purposes. In this new model, college supervisors were being asked to do more than observe but rather mentor alongside the cooperating teacher. In these 3-way conversations, student teachers were able to access the thought-processes and wisdom of both mentors. The goal was to create a growth mindset in candidates where the norm became “How can I learn from my mistakes and become a better teacher?” (Brondyk and Cook, 2020).

Training mentors to use the tool

Preparing mentors to enact these new ways of working together was a critical component of implementing the new tool, as PEC members believed that it was important that all participants understand the theory behind the tool and be equipped with the knowledge and skills to fully participate in the developmental process. To accomplish this, program leaders created unique learning

opportunities for each group of mentors, since prior to this there had been no formal training for mentors. Participation in the training sessions was high with attendance rates consistently over 83.3%. Training involved cooperating teachers coming to campus for a one-time workshop that was followed by monthly, online modules spread over the course of the semester. College supervisor meetings were extended to monthly and included in-depth descriptions of STAT, with a detailed description of each item, examples of “meeting expectations” and suggestions of way that mentors might use the tool to promote growth in their student teachers. The idea was to help mentors understand the vision while also providing them with tools to implement it.

The Student Teaching Assessment Tool became central to the deep-sea change in this program. What began as minor changes to a tool based on accreditation feedback, became a visioning process that led to a stronger sense of program identity.

Conceptual framework

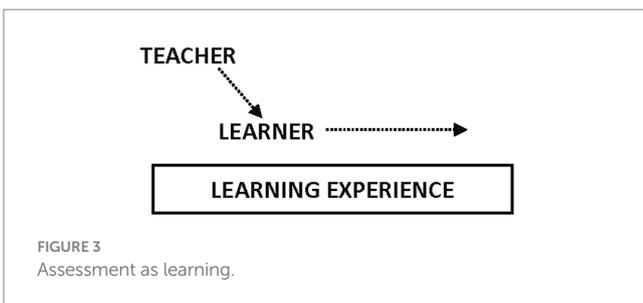
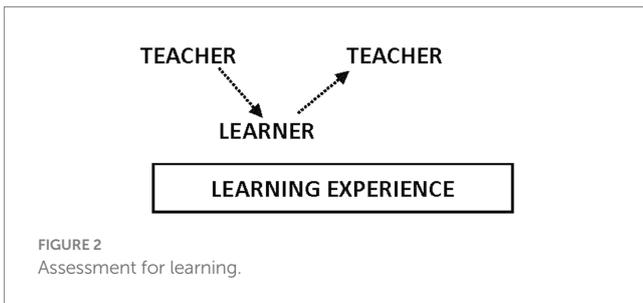
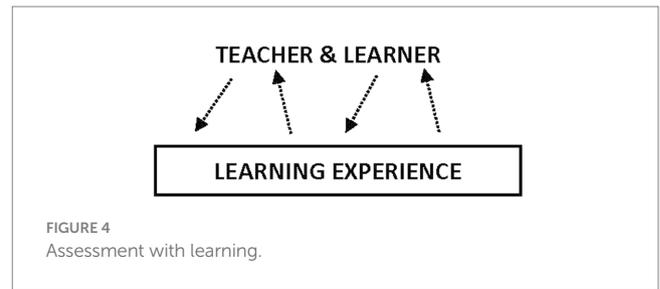
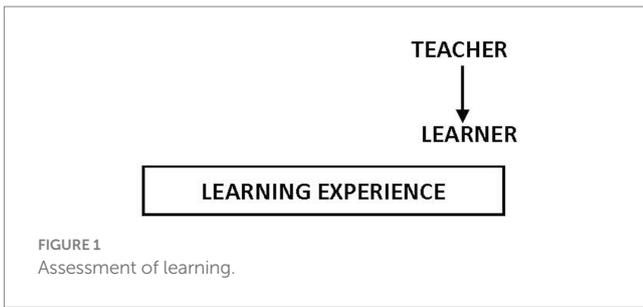
Assessing student learning is central to teaching (Stolz, 2017). There is currently a strong emphasis on data-driven instruction for P-12 teachers (Lee et al., 2012), who are expected to be able to analyze and use student test data to inform their teaching. As a result, teachers spend considerable time testing students and analyzing test results, yet in most cases, students rarely spend time with these assessments beyond taking the test and looking at their grade.

Some researchers propose that we think differently about student assessment (Chappuis, 2014). Wiggins and McTighe (2005) suggested that assessments should be used as opportunities for students to learn. Rather than merely using assessments to give a grade at the end of a learning experience, students should be given access to their assessments so that they can learn from their mistakes and master the material. This view of grading shifts the emphasis from evaluating to assessing, from teacher to student.

Assessments are equally important in teacher education as they help both programs and instructors determine what students know and are able to do in terms of learning to teach. Several nationally-known assessments are used by many educator preparation programs [e.g., (Danielson, 2011) *A Framework for Teaching*]. These types of evaluation rubrics remain the most prevalent way to assess student teachers (Sandholtz and Shea, 2015), but are not often used as opportunities to learn as student teachers are not always involved in the evaluation process. Contributing to this disconnect is a recent trend in some states requiring all EPPs to use a common assessment.

Many school districts also use these commercially-produced instruments in their annual teacher evaluation process. Administrators mark evaluations based on teaching observations which then contributes to the teachers' overall effectiveness ratings. Some, but not all, principals use this as an opportunity to discuss growth areas and ways to support teachers, especially during their induction years (Derrington and Campbell, 2018).

The shift from thinking of assessment as merely evaluative provides a framework for this study. Assessment of learning is often equated with summative assessments which tend to be administered toward the end of the learning experience. These are used primarily as a means for the teacher to evaluate the learner, who are basically told what they need to improve (Figure 1).



student work together throughout the learning experience in order to support the learner’s growth and development. Teachers benefit by garnering concrete data regarding what the learner knows and still needs to learn. Rather than being told what needs to improve, learners are assisted in both nominating growth areas and identifying actionable steps to improvement. This creates an interactive, iterative process where the learner is supported in interpreting the data and collaboratively working to develop their practice with the help of an experienced educator (Figure 4).

Considering how much time students spend being assessed, it is critical to rethink their role in the learning process. This shift in thinking brought about questions in this EPP regarding assessments, how they are used and by whom.

Materials and methods

In this mixed-methods study, I examined how mentors in one small, liberal arts college used the program’s uniquely-designed assessment tool to promote growth and development during student teaching. Research questions were: How do tools contribute to the identity of a program question? What value does using an internally-designed evaluation tool add to an educator preparation program?

Participants

Participants were chosen from mentors who oversee student teaching, both college supervisors and cooperating teachers. Eight college supervisors were identified because they had mentored candidates in each of the seven semesters of the study. Of the 246 K-12 cooperating teachers that hosted student teachers, volunteers were randomly selected from the pool of cooperating teachers that worked with corresponding college supervisors, as 3-way conversations were a critical part of the data. Qualtrics results of STAT were analyzed for all 246 cooperating teachers.

Participation was strictly voluntary and participants were able to withdraw at any time. Pseudonyms were used for each student teacher, cooperating teacher and college supervisor and were used in all presentations and written materials regarding the results of this study.

Even though the majority of data collected in this study were materials that the participants would normally submit in the course of their work with the Education Department, all audio recordings, transcripts and field notes were kept in a locked file cabinet in my locked office and will be destroyed at the completion of the study.

The greatest potential risk in this study was to the student teachers, who are receiving a grade for their student teaching experience and

William (2011) added to our understanding of assessment with his description of assessment *for* learning, which equates to formative assessment as evaluations typically occur throughout the learning experience and serve to inform the instructor’s teaching. In this model, learners are not explicitly informed of the results as the information tends to be more for teachers as the tool informs them what students know and still need to learn, which then impacts future lessons (Figure 2).

In recent years, assessment *as* learning emerged as yet another way to think of assessment. In this model, assessment data is shared with students who then spend time reflecting on what they know and still need to learn throughout the learning experience. In doing so, they take ownership of their learning in ways that do not occur when assessments are used to merely evaluate or inform the teacher. Standards-based grading (Sadik, 2011; Vatterott, 2015) is a practice that reflects this use of assessments as I Can students break down standards into learning targets that students can monitor on their own (Figure 3).

While helpful in some respects, these descriptors of assessments seem somewhat simplistic. Conceptually they do not adequately capture the collaborative, developmental nature of assessment use in some programs. Launching from this framework, a fourth descriptor seems plausible: assessment *with* learning. In this model, teacher and

TABLE 1 Ability 1 of STAT: ethical educator CVI (0.33 overall) (0.81).

Disposition	Essential			Useful			Not necessary		
	CD	A	T	CD	A	T	CD	A	T
Demonstrates responsibility and maturity	10	7	8	2	3	0	0.56	0.62	0.75
							0.83	0.7	1.0
Demonstrates respect	10	7	7	2	3	1	0.83	0.7	0.88
Displays a positive attitude when interacting with students	12	10	7	0	0	1	1.0	1.0	0.88
Demonstrates a commitment to reach all students	11	10	7	1	0	1	0.92	1.0	0.88
Demonstrates personal integrity	11	9	8	1	1	0	0.92	0.9	1.0
Demonstrates equity	8	6	7	4	4	1	0.67	0.6	0.88
Demonstrates enthusiasm for the content	5	4	4	7	6	4	0.42	0.4	0.5
Demonstrates passion for teaching	11	8	7	1	2	1	0.92	0.8	0.88
Demonstrates perseverance	9	8	8	3	2	0	0.75	0.8	1.0
							0.82	0.76	0.84

this grade was determined by both the college supervisor and cooperating teacher and the Director of Student Teaching was also one of the researchers in this study. To minimize this risk to student teachers, they were asked to complete the survey only after grades have been submitted so that they felt free to decline to participate and do so freely without feeling that it will in any way affect their grade.

Likewise, the college supervisors were employees of the college and because of this may have felt pressure to participate. We also recognize that people can feel vulnerable when they open their practice to others (i.e., being audiotaped having a mentoring conversation). We made it clear both verbally and in writing (consent form) that the college supervisors' participation was strictly voluntary and would in no way affect their employment status. We also tried to reassure the supervisors that we were analyzing the assessment tool and the growth of the student teacher, not their mentoring practice.

Data collection

I examined seven semesters of data including Qualtrics survey results and Working Versions of STAT that mentors used to document observations, conversations, suggestions, and growth. To establish content validity of the tool, I conducted three separate reviews with area curriculum directors, members of our Teacher Education Council (TEC), and our Alumni Network (HEAN). TEC is a partnership with area teachers and principals that meets once a semester, while HEAN is a group of alumni education leaders throughout our state. Table 1 shows the Lawshe coefficients from each of the three different groups for one of the abilities (Ethical Educator) as an example. The question that I ask with each individual disposition is, "Is this disposition considered essential, useful but not essential, or not necessary?" The number in the "essential" and "necessary" columns were the actual number of group members that selected that choice, while the "not necessary" column shows the content validity index for each of the groups. Based on this work, I determined that Enthusiasm for Content was not a disposition and so was removed from the list of dispositions and placed in the Effective Instructor Ability, because the department still considered it an important indicator. That I needed to include in our assessment.

TABLE 2 Cronbach alpha-cooperating teachers' ratings on STAT.

Cronbach alpha	Cronbach alpha based on standardized items	N of items
Spring 20xx		
0.793	0.754	20
Fall 20xx		
0.631	0.591	15

Inter-rater reliability (IRR) on STAT has been established in several ways. First, EPP faculty established IRR of at least 80% (i.e., social science statistical standard) in the Ethical Educator section of STAT by discussion and then individual ratings of scenarios designed for each element. Faculty discussed their ratings and came to consensus on how each element would be rated for in-class dispositional ratings.

Cooperating teachers

Internal consistency estimates of reliability were also computed for STAT. Final ratings for student teachers who had two or more cooperating teachers were compared for consistency. Values for the coefficient alpha (Cronbach's Alpha) are indicated on Table 2.

Each of these coefficient alpha scores indicates reliability levels below the 80% mark. Although many of the cooperating teachers had participated in our on-campus training sessions that included a significant portion devoted to reviewing STAT, the lower coefficient alpha scores point to the need for further work to establish internal consistency.

College supervisors

Finally, inter-rater reliability was established with the cohort of college supervisors. During an entire academic year, program leaders met monthly with the supervisors for either a three or one-hour session and at least half of each meeting was devoted to establishing inter-rater reliability (see Table 3). One STAT Ability was addressed at each meeting using the following process: (a) discussing our respective understandings of each individual item, including what it would look

TABLE 3 Inter-rater reliability college supervisors.

Date	Ability	Format
August 20xx	All	Overview of STAT
September 20xx	Skilled communicator	Video
October 20xx	Decision maker	Video
November 20xx	Ethical educator	Scenarios
December 20xx	Curriculum developer	Scenarios
January 20xx	Effective instructor	Video
February 20xx	Engaged professional	Scenarios
March 20xx	Reflective practitioner	Scenarios
November 20xx	All	Video

TABLE 4 Inter-rater reliability percentage for college supervisors on STAT.

Date	N of CS	Numerical average	Percentage
11/13/20xx	9	7.2	80%

and sound like in the classroom, (b) analyzing the language in the proficiency levels and discussing the relative differences between each level (e.g., What is the difference between “meets” and “developing” expectations?), and (c) engaging in some sort of inter-rater reliability activity. If large discrepancies on any items were discovered, these were discussed at the following meeting. During one meeting, the group spent the entire hour establishing inter-rater reliability by watching a video clip of a teacher candidate teaching a lesson and then rating the lesson on STAT. Each supervisor completed his/her version of STAT and then the numbers were calculated resulting in a numerical score of 7.2 or 80%.

Additional qualitative data were collected as part of the professional development offered to all mentors, including recordings of monthly meetings, responses to prompts, and semi-structured interviews. Quantitative data came from running *t*-tests and determining means and percentages for individual items, abilities, semesters and subcategories of participants (Table 4).

Data analysis

Quantitatively, the STAT data was tested for significance using SPSS. After determining the means and standard deviation, a *t*-Test: Paired Two Sample for Means was run on STAT data (0.038) to establish Pearson's *r* value (0.83) as means to analyze the two sets of scores (S1 and S7). In addition, the frequency statistics (e.g., which items/abilities have the most exceeds?) were determined. Data analysis of the qualitative data included developing an initial set of analytical categories based on: (a) a review of the literature, (b) the researchers' experience of and knowledge about experiential learning, and (c) the goals of the study. Then data was coded based on these categories. Data were coded twice. The initial round looked for means of assisting performance which included the various moves that mentors make to assist beginning teachers as they move from assistance to independence. Codes included: M = Modeling; CM = Contingency Management; F = Feeding-back; I = Instructing (linguistic);

Q = Questioning (linguistic); Q-assess (recitation); Q-assist; CS = Cognitive-structuring; Structure of Explanation (CS-E); and Structures for Cognitive Activity (CS-CA). A second set of codes were applied that analyzed the developmental nature of responses: G = Evidence of Growth; M = Evidence of Mentoring; MM = Mentoring Move; DL = Use of Developmental Language. In addition, analysis of the focal items identified by triads and ways that mentors worked together to support the candidate's growth included listening to audio-recordings of 3-way conversations.

The analysis process was multi-stepped, going back and forth between the quantitative and qualitative data. Data sets were triangulated to look for significant patterns. The “Exceeds Expectations” category emerged as interesting and worthy of further analysis. Analysis included calculating the frequency for individual student teachers and then coding the reasons provided by the mentors looking for elements that were relational, dispositional, and growth identifying/producing. Researchers also correlated evidence of growth with evidence of mentoring in the 3-way conversations and both versions of STAT (Qualtrics and Working Version).

Limitations of the study

The primary limitation of this study is replicability as the study analyzes one assessment tool that was developed and exclusively used by the program. This limits its generalizability and external validity, which restricts the ability to draw broad conclusions about its effectiveness or applicability in other settings or populations. Another concern is the potential for bias, as the creators of the tool may have a vested interest in demonstrating its success, potentially compromising the objectivity of the study.

Results

Both sets of data reveal that the mentors in this program adopted a more developmental stance. The STAT mean analysis shows that both sets of mentors used more 2's (developing) and less 4 ratings in the seventh semester than in the first. The means got lower (smaller) over time with less 4's and more 3's and 2's (Table 5).

Table 1 shows that there was an increase in percentage of 2's from S1 (3.07) to S7 (6.78) with a slight increase in percentage of 3's from S1 (78.28) to S7 (81.39). The mean for college supervisors decreased steadily over time from S1 (3) to S7 (2.98) with the mean for supervisors decreasing more than the mean for cooperating teachers from S1 (3.1) to S7 (3.06). Of particular notes is the fact that the percentage of 4's for college supervisors significantly decreased from S1 (9.89) to S7 (4.84) and the percentage of 2's for increased from S1 (4.12) to S7 (5.06).

Another data point came from the program's monitoring system, which tracks cohort scores. The cohorts from semester seven received an average score of 3.2, unlike previous semesters in which the majority of candidates were rated consistently closer to 4. This shows that “Meets Expectations” had become more of the norm over time.

Qualitative data confirms the trend toward a more developmental stance. In meeting transcripts, mentors can be heard using developmental terminology in which they talk about learning to teach as a life-long endeavor as there are always ways to improve. For

TABLE 5 Mentor means.

	Semester	# STS	Rating	All	%	Overall mean	CS	%	CS mean	CT	%	CT mean
S1	F16	35	4	716	16.91	3.08	185	9.89	3	531	22.45	3.1
			3	3,315	78.28		1,573	84.12		1742	73.66	
			2	130	3.07		77	4.12		53	2.24	
S2	S17	46	4	560	10.07	2.96	181	7.45	2.95	379	12.1	2.96
			3	4,484	80.62		2082	85.68		2,402	76.69	
			2	365	6.56		111	4.57		254	8.11	
S3	F17	38	4	559	12.39	3.08	193	9.23	3.06	373	15.07	3.1
			3	3,836	85.06		1853	88.66		2030	82.02	
			2	81	1.8		33	1.58		49	1.98	
S4	S18	57	4	823	11.55	3.03	173	5.82	2.99	650	15.63	3.05
			3	5,917	83.01		2,673	90		3,244	78.02	
			2	270	3.79		86	2.9		184	4.43	
S5	F18	21	4	315	12.15	3.05	85	7.15	2.98	230	16.38	3.1
			3	2,151	66.67		1,040	87.54		1,111	79.13	
			2	92	3.55		40	3.37		52	3.7	
S6	S19	48	4	624	9.71	3	90	3.62	2.94	533	13.52	3.04
			3	5,411	84.2		2,242	90.26		3,171	80.44	
			2	268	4.17		107	4.31		160	4.06	
S7	F19	26	4	346	10.86	3.01	68	4.84	2.98	239	16.39	3.06
			3	2,593	81.39		1,256	89.46		1,105	75.79	
			2	216	6.78		71	5.06		95	6.52	

S1 = Semester 1; S2 = Semester 2; S3 = Semester 3; S4 = Semester 4; S5 = Semester 5; S6 = Semester 6; S7 = Semester 7. F16 = Fall 2016; S17 = Spring 2017; F17 = Fall 2017; S18 = Spring 2018; F18 = Fall 2018; S19 = Spring 2019; F20 = Fall 2020.

example, a 4th grade cooperating teacher analyzed informational essays with her student teacher. As they sorted the essays, the cooperating teacher modeled being a learner when she realized that they were looking at too many objectives and had to narrow their lens. She explicitly talked about how good teachers constantly need to learn from new situations in order to grow professionally. Similarly, college supervisors began using developmental language to help promote growth by identifying strengths and targeting growth areas. Supervisors who would have previously named the problem and told the student teacher how to fix it, now began asking questions to get them to reflect and weigh options for future lessons. One supervisor said, “I am much more intentional. I talk with them at length about “why” they are doing what they are doing” (CS2), while another expressed, “I’ve been more conscious of incremental growth, even when the student teacher is awesome” (CS4). In 3-way conversations, supervisors were heard talking explicitly with student teachers about their developmental growth, using STAT to plan ways to support the candidate. The following quotes are from survey data in which college supervisors talk about using STAT developmentally:

The instrument itself has made a big difference to me because we can get caught up in generalities and that’s not the best way to help a new young teacher. You have to lay out specifics that are research based. It’s not my opinion and I’ve seen a lot of good teaching over the years, but sometimes you need to quantify things and be specific. And if you are going to offer a methodology and suggestions then they should be based on successful practice.

And that’s what the tool drives us to take into account. I think that I am much more effective in my role because I have the instrument and I have the team that I can work with effectively because we are all talking from the same sheet of music (CS3).

I often focus on the green (instruction) items on STAT. When they realize that delivery of instruction is key to building understanding. Are they assessing constantly as they judge responses, evaluating progress on assignments, etc.? My statement is this—you are moving from the act of teaching to the art of teaching. The rubric shows this growth and students celebrate the shift and at times note, “I can feel it” (CS2).

I am now more mindful, deliberate and intentional with my conversations. STAT helps us to focus on key topics and is a great guide for difficult situations (CS8).

The shift to a more developmental stance was also visible in the student teachers. They tend to be strong students used to receiving good grades, so hearing statements about data and growth areas was a real change. This new way of thinking is evident in the following examples:

I was pleasantly surprised at how well that lesson went! I think that one of my strongest areas within teaching these lessons was classroom management. Based on our previous 3-way conversation, I was really conscious of the volume of my voice and kept it a bit softer so students would pay more close attention.

I think I can take that STAT item off my list. If I could do anything differently with this lesson for next time, I would most definitely try to budget my time better in advance (ST7).

Overall, through this semester I have been able to reflect on the positives, negatives, student learning, personal learning, and create goals for the future. The unique part of teaching is that it is different every time it is done. I can teach a lesson one hour and the hour later, and I could have completely different experiences. Teaching is a profession of constant evaluation and adjustment, and that flexibility and adaptability are some things I look forward to practicing. I value experiences like this to be able to think about my own actions and what I can do best for the benefit of the class and each student's learning. I hope to continue to keep learning, even long after classes have finished. I will hold onto the tools and strategies I have gained to continue reflecting and setting goals down the road in my teaching career (ST3).

All of the data points to a shift from having an evaluative mindset focused on grades and evaluating students at the end of student teaching to more of an assessment mindset focused on improving practice throughout the learning experience.

Discussion

Learning to teach requires a complex combination of knowing content, pedagogical skills and how to put those all together to meet the needs of learners. Pedagogical content knowledge (Shulman, 1987) describes this unique form of knowledge that teachers must possess and execute in order to be successful. It stands to reason that learning a practice that is so complex requires assessments that are equally nuanced.

The Student Teaching Assessment Tool was created to ground mentors and student teachers in a new way of working together that ultimately reflected the values of the program. The “with”-ness of the assessment process, in which mentors and student teachers used data to continuously work on growth areas, became an integral part of the program's identity. Three-way conversations based around STAT took on a developmental tenor. The collaborative elements of the tool influenced the ways that mentors worked with candidates and led to more targeted and constructive mentoring. Use of the tool, combined with extensive professional development, resulted in changed attitudes about the developmental nature of learning to teach.

The accreditation process encourages programs to continually improve. This involves constantly analyzing data, reflecting on outcomes, listening to feedback, identifying weak areas and working to improve them. In this way each program has its own process and products that make it unique. Although programs are guided by state-mandated standards, the continuous improvement process has the potential to lead to distinctive features. Discounting the significance of this process by requiring programs to utilize mandated tools has the potential to lessen the quality of programs because the incentive to have ongoing conversations diminishes.

Identity is critical to programs as it reflects what they believe about learning and specifically how candidates learn to teach. Activities that revolve around identity can energize and unify a program. In this case, the multi-year experience of designing a new

tool led to a clearer articulation of beliefs that created more coherence across the program and gave concrete direction to mentors. The ongoing conversations about items on the tool, clarity of language, and distinctions between proficiency levels all led to more consistency. Mentors and faculty began using consistent language and practices that reinforced the developmental nature of learning to teach. Having a strong identity that was clearly articulated provided guidance and directions to mentors that allowed them to understand and adhere to the goals of the program. Without these unifying structures, mentors would have been left on their own to do as they see fit.

Identity also reflects what programs hope to project to broader constituencies, including prospective students. Program-wide ideas and materials, like assessment tools, mirror who we are and what we believe about teaching and learning. They also act as selling points, drawing students to programs, which is critical at this time when teacher preparation is being called into question and numbers are decreasing. The adoption of a clear vision of learning to teach, with its accompanying materials and training, point to the value-added by engaging in this type of introspective process.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Participants gave permission for only the researcher/author to access the interview and recording transcripts. Requests to access these datasets should be directed to brondyk@hope.edu.

Ethics statement

The studies involving humans were approved by Hope College Human Subjects Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

SB: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Brondyk, S. (2020). Mentoring as loose coupling: theory in action (Invited Chapter). *The Wiley- Blackwell International Handbook of Mentoring: Paradigms, Practices, Programs, and Possibilities*, Eds. Irby, B., Boswell, J., Searby, L., Kochan, F., and Garza, R. Hoboken, NJ: John Wiley & Sons, Inc.
- Brondyk, S., and Cook, N. (2020). How one educator preparation program reinvented student teaching: a story of transformation, New York, NY: Peter Lang Publishing.
- Burden, S., Topping, A. E., and O'Halloran, C. (2018). Mentor judgements and decision-making in the assessment of student nurse competence in practice: a mixed-methods study. *J. Adv. Nurs.* 74, 1078–1089. doi: 10.1111/jan.13508
- Chappuis, J. (2014). Thoughtful assessment with the learner in mind. *Educ. Leadersh.* 71, 20–26.
- Danielson, C. (2011). *Enhancing professional practice: a framework for teaching*, Alexandria, VA: ASCD.
- Derrington, M. L., and Campbell, J. W. (2018). High-stakes teacher evaluation policy: US principals' perspectives and variations in practice. *Teach. Teach.* 24, 246–262. doi: 10.1080/13540602.2017.1421164
- Feiman-Nemser, S. (2001). From preparation to practice: designing a continuum to strengthen and sustain teaching. *Teach. Coll. Rec.* 103, 1013–1055. doi: 10.1111/0161-4681.00141
- Lee, M., Louis, K. S., and Anderson, S. (2012). Local education authorities and student learning: the effects of policies and practices. *Sch. Eff. Sch. Improv.* 23, 133–158. doi: 10.1080/09243453.2011.652125
- Lejonberg, E., Elstad, E., and Christophersen, K. A. (2015). Mentor education: challenging mentors' beliefs about mentoring. *Int. J. Mentor. Coach. Educ.* 4, 142–158. doi: 10.1108/IJMCE-10-2014-0034
- Sadik, A. M. (2011). A standards-based grading and reporting tool for faculty: design and implications. *J. Educ. Technol.* 8, 46–63.
- Sandholtz, J. H., and Shea, L. M. (2015). Examining the extremes: high and low performance on a teaching performance assessment for licensure. *Teach. Educ. Q.* 42, 17–42.
- Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harv. Educ. Rev.* 57, 1–23. doi: 10.17763/haer.57.1.j463w79r56455411
- Stolz, S. (2017). Can educationally significant learning be assessed? *Educ. Philos. Theory* 49, 379–390. doi: 10.1080/00131857.2015.1048664
- Ulvik, M., and Sunde, E. (2013). The impact of mentor education: does mentor education matter? *Prof. Dev. Educ.* 39, 754–770. doi: 10.1080/19415257.2012.754783
- Vatterott, C. (2015). *Rethinking grading: meaningful assessment for standards-based learning*. Alexandria, VA: ASCD.
- Wiggins, G. P., and McTighe, J. (2005). *Understanding by design*. Alexandria, VA: ASCD.
- William, D. (2011). What is assessment for learning? *Stud. Educ. Eval.* 37, 3–14. doi: 10.1016/j.stueduc.2011.03.001