

SPECIAL ISSUE
**Navigating the Contested Terrain of
Teacher Education Policy and Practice**

education policy analysis
archives

A peer-reviewed, independent,
open access, multilingual journal



Arizona State University

Volume 26 Number 35

March 12, 2018

ISSN 1068-2341

Haberman Star Teacher Interview as a Predictor of Success in Urban Teacher Preparation

Jennifer H. Waddell



Jacob M. Marszalek

University of Missouri–Kansas City
United States

Citation: Waddell, J. H., & Marszalek, J. M. (2018). Haberman Star Teacher Interview as a predictor of success in urban teacher preparation. *Education Policy Analysis Archives*, 26(35).

<http://dx.doi.org/10.14507/epaa.26.2808> This article is part of the Special Issue, *Navigating the Contested Terrain of Teacher Education Policy and Practice*, guest edited by Elena Aydarova and David Berliner.

Abstract: In urban schools, along with skills for effective teaching, successful teachers must also possess values and belief systems conducive to teaching effectively in diverse settings (Becker, Kennedy, & Hundersmarck, 2003; Haberman, 2008; Metzgar & Wu, 2008). As demonstrated in CAEP standard 3, there is a critical need for EPPs to admit candidates who have both the dispositions to be effective teachers in urban schools and the propensity for success within the preparation program. The Haberman Star Teacher Interview is a commercial teacher selection instrument designed for use in selecting teachers for urban schools. This study examines the validity of the instrument as a

Journal website: <http://epaa.asu.edu/ojs/>

Facebook: /EPAAA

Twitter: @epaa_aape

Manuscript received: 11/1/2016

Revisions received: 1/22/2017

Accepted: 2/17/2017

selection instrument for teacher preparation programs. The selection instrument was administered to 109 students before entry into an urban teacher preparation program at an urban university in the U.S. Midwest. Inter-rater agreement and principle components analysis provided evidence of reliability and structural validity of the multi-part Haberman scores. Logistic regression analyses supported the validity of using the Haberman scores to predict later program attrition, but not in the manner recommended by its developers. Within this paper, the authors recommend the cautious use of the instrument in urban teacher preparation. Application of scoring and program implications are discussed.

Keywords: urban teacher education; teacher education preparation; recruitment and retention; selectivity; quantitative research; preservice teacher preparation

Entrevista con Haberman Star Teacher como un predictor de éxito en la preparación de docentes urbanos

Resumen: En las escuelas urbanas, junto con las habilidades para una enseñanza eficaz, los docentes exitosos también deben poseer valores y sistemas de creencias que propicien una enseñanza eficaz en entornos diversos (Becker, Kennedy & Hundersmarck, 2003; Haberman, 2008; Metzgar & Wu, 2008). Como se demuestra en la norma CAEP 3, existe una necesidad crítica de que los EPP admitan candidatos que tengan las disposiciones para ser docentes eficaces en las escuelas urbanas y la propensión al éxito dentro del programa de preparación. La Entrevista Haberman Star Teacher es un instrumento comercial de selección de docentes diseñado para utilizar en la selección de docentes de las escuelas urbanas. Este estudio examina la validez del instrumento como un instrumento de selección para los programas de preparación docente. El instrumento de selección fue administrado a 109 estudiantes antes de ingresar a un programa de preparación de docentes urbanos en una universidad urbana en el medio oeste de Estados Unidos. El acuerdo entre evaluadores y el análisis de componentes principales proporcionaron evidencia de la confiabilidad y la validez estructural de los puntajes de Haberman en varias partes. Los análisis de regresión logística respaldaron la validez del uso de los puntajes Haberman para predecir el desgaste posterior del programa, pero no de la manera recomendada por sus desarrolladores. En este documento, los autores recomiendan el uso cauteloso del instrumento en la preparación de docentes urbanos. Aplicación de puntuación y las implicaciones del programa se discuten.

Palabras clave: educación de docentes urbanos; preparación de la educación docente; reclutamiento y retención; selectividad; investigación cuantitativa

Entrevista com Haberman Star Teacher como preditor de sucesso na preparação de professores urbanos

Resumo: Nas escolas urbanas, juntamente com as habilidades para o ensino efetivo, os professores bem-sucedidos também devem possuir valores e sistemas de crenças que promovam o ensino efetivo em diversos ambientes (Becker, Kennedy & Hundersmarck, 2003; Haberman, 2008; Metzgar & Wu, 2008). Conforme demonstrado pelo padrão do CAEP 3, há uma necessidade crítica de PEPs admitir candidatos que tenham as provisões para serem professores eficazes em escolas urbanas e a propensão a ter sucesso dentro do programa de preparação. O Haberman Star Teacher Interview é uma ferramenta de seleção de professores comercial projetada para ser usada na seleção de professores em escolas urbanas. Este estudo examina a validade do instrumento como uma ferramenta de seleção para programas de preparação de professores. A ferramenta de seleção foi administrada a 109 alunos antes de entrar em um programa de treinamento de professores urbanos em

uma universidade urbana no Centro-Oeste dos Estados Unidos. O acordo entre avaliadores e a análise de componentes principais evidenciou a confiabilidade e validade estrutural dos escores de Haberman em várias partes. As análises de regressão logística apoiaram a validade do uso dos escores de Haberman para prever o desgaste posterior do programa, mas não da maneira recomendada pelos desenvolvedores. Neste documento, os autores recomendam o uso cauteloso do instrumento na preparação de professores urbanos. A aplicação do escore e as implicações do programa são discutidas.

Palavras-chave: educação urbana de professores; preparação da formação de professores; recrutamento e retenção; seletividade; pesquisa quantitativa

Introduction

“When teacher education programs make decisions about [whom] to accept into their program, they are essentially making predictions and gambling on who they believe will grow and develop into highly effective teachers” (Clark, 2010, p.1). As they respond to the challenge of preparing teachers to work effectively in all communities, teacher education programs are facing increased demands for accountability regarding candidate performance and selection (Hollins, 2012; Sandholtz & Shea, 2012). Universities and colleges of education are not only being scrutinized for the performance of their graduates, but also asked to provide evidence that candidates are demonstrating that they possess the qualities of an effective teacher. Furthermore, these Educator Preparation Programs (EPPs) are now being evaluated on the promotion, retention, and effectiveness of their graduates (Council for the Accreditation of Educator Preparation [CAEP], 2017).

Meanwhile, among the current policy and practice debates in teacher education is the selection of candidates. CAEP Standard 3 focuses on Candidate Quality, Recruitment and Selectivity. Within this standard, CAEP (2016a) calls for EPPs to enact “admissions selectivity that builds an able and diverse pool of candidates” (p. 1), and charges EPPs with “the responsibility to recruit a diverse candidate pool that mirrors the demography of the student population served” (p. 2). Concurrently, CAEP (2016b) has mandated minimum standards defining “high academic achievement and ability” (p. 1) for candidates within a teacher preparation program. However, in 2012, the American Association of Colleges of Teacher Education (AACTE) asserted that, “Course grades, grade point averages (GPAs), and scores on college entrance exams are not good predictors of candidates’ ability to acquire the multiple specific skills that teaching requires” (AACTE, 2012, p. 1). In urban schools, along with skills for effective teaching, successful teachers must also possess values and belief systems conducive to teaching effectively in diverse settings (Becker, Kennedy, & Hundersmarck, 2003; Haberman, 2005a; Metzgar & Wu, 2008). In addition, Berliner (2005) illuminated the difficulty of “testing for teacher quality” in arguing that many characteristics of effective teachers, such as evidence of student learning and “logical, psychological and moral dimensions of teaching” (p. 208), cannot be assessed through paper and pencil tests, and therefore, are costly and/or impossible to measure prior to employment. This difficulty increases when considering selecting candidates for entry into a teacher preparation program. It is not surprising, then, that CAEP Standard 3 has ignited the policy debate regarding what constitutes an effective teacher and on what criteria EPPs should base admissions decisions.

Although CAEP (2016b) contends that EPPs can determine “whether the CAEP minimum criteria [for academic achievement and ability] will be measured (1) at admissions, OR (2) at some other time prior to candidate completion” (p. 1), there is a critical need for EPPs to admit candidates

who have both the dispositions to be effective teachers in urban schools and the propensity for success within the preparation program. In a position statement, AACTE (2012) supported selectivity in teacher education by asserting that:

Candidates entering educator preparation should possess adequate literacy, quantitative, and reasoning skills to be able to assimilate their preparation and to serve as a positive model to PK-12 students. Further, candidates should have an affirmative disposition and desire to advance the learning of all students—including those faced with the most challenging circumstances—so as to support students' mastery of the knowledge and skills required of their age and educational level (p. 1).

The position paper also indicated that many skills must be honed while candidates are in teacher education programs, and “prior estimates of these skills have not been found to be reliable” (AACTE, 2012, p. 2). Yet, annually, candidates enter teacher education programs and are not able to develop the skills needed to be effective teachers. So, while both CAEP and AACTE agree that EPPs must increase admissions standards and selectivity, there is a question as to how to better screen for non-academic skills and dispositions. In addition, there is a need to determine whether or not candidates will successfully complete the EPP, as well as secure and be retained in teaching positions once they graduate.

It is pertinent, then, for EPPs to analyze the success and job placement of teacher candidates as they make decisions regarding selection, training and promotion of new candidates. In reflecting on the current demands from the literature, and the policy and practice implications of CAEP, the authors began a review of their program and graduate success. Within the EPP involved in this study, 96% of program completers secured teaching positions upon graduation. While national data on job placement of beginning teachers is not widely known, published data from EPPs and/or states report teacher education graduate job placements rates as 57-74% in North Carolina (Bastian, n.d), 61% for Central Michigan (CMU, 2006) and 83% for the University of Kansas (University Career Center, 2012). Therefore, the job placement of teachers graduating from the EPP in this study is exceptionally high. Retention rates for graduates of the EPP is also comparatively high. According to the National Center for Education Statistics (2015), only 82% of teachers are still teaching after five years. Other studies have shown even grimmer data, demonstrating that 40-50% of beginning teachers leave the profession in the first five years (Ingersoll, 2003). Data also show that “high-poverty, high-minority, urban, and rural public schools have among the highest rates of turnover” (Ingersoll, Merrill, & Stuckey, 2014, p. 23). However, within the EPP involved in this study, 96% of graduates who began teaching were still teaching beyond their third year. To date, the five-year urban teacher retention rate for all program graduates is 92%. Given the high success rates of graduates in the program, the authors assert that graduates of the program have an increased likelihood of being employed and retained in the teaching profession. Therefore, the authors began looking into their practices to determine if lessons and policy implications could be gleaned from their work. Two areas of particular focus were the admissions procedures used in the program and the field-based design of the program. The authors hypothesized that success in the program equates to success in the field, and, due to the current demands for selectivity in teacher education, sought to determine if components of their selection criteria could predict success in the program, and, therefore, in teaching. Thus, the purpose of the study was to explore the validity of using an existing commercial teacher selection instrument that was used as a selection instrument for the EPP.

Perspective

Hollins (2012) reminds us that urban schools face a unique set of challenges, resources, and opportunities, and that effective teaching in those settings demands careful attention to the ways in which we situate teacher preparation for that unique endeavor. Higher education has been accused of not preparing teacher candidates to face the challenges of urban schools by focusing on the status quo rather than the changes needed for reform (Doll, 2009). Likewise, the U.S. Department of Education (2000) has cited teacher preparation as a barrier to improving education in the United States, resulting in over a decade of calls for teacher preparation institutions to redesign programs to better fit the changing demographics extant in the nation's schools.

Reexamining EPPs

EPPs proposing to prepare teachers for urban schools need to make significant changes to their content and structure to include: (a) curricular changes, including a focus on diversity and culturally responsive practice, knowledge of self and others, and an emphasis on connections from theory to practice (Carpenter-LaGattuta, 2004; Cochran-Smith, 2004; Darling-Hammond, 2006, 2010; Darling-Hammond & Baratz-Snowden, 2007; Gay, 2000, 2004; Ladson-Billings, 1994, 2006; Milner, 2010); (b) a focus on the preparation for the realities of urban communities with an emphasis on social justice (Cochran-Smith, 2004; Gay, 2004; Hollins, 2012; Ladson-Billings, 1994, 2001; Weiner, 1999; Zeichner, 2003); (c) extended and closely supervised field experiences in urban schools, which may include coursework taught in urban classrooms and communities (AACTE, 2013; Conaway, Browning & Purdum-Cassidy, 2007; Darling-Hammond, 2010; Darling-Hammond & Baratz-Snowden, 2007; Ladson-Billings, 1994, 2006; National Council for Accreditation of Teacher Education, 2010; Zeichner, 2003, 2010, 2011); and (d) the recruitment of a diverse pool of candidates specifically for teaching in urban schools (AACTE, 2013; Darling-Hammond & Baratz-Snowden, 2007; Haberman, 2005a; Ladson-Billings, 1995; Ryan & Alcock, 2002; Tredway, 1999; Weiner, 2000). The focus on teacher education reform, accountability for the performance of graduates, and the necessity for robust selectivity in teacher education demands that EPPs examine selection procedures to ensure that they are admitting students with the potential for effective teaching.

Selection of Candidates

In urban schools, along with skills for effective teaching, successful teachers must also possess values and belief systems conducive to teaching effectively in diverse settings. In a paper presented at the 2003 American Educational Research Association annual meeting, Becker, Kennedy, and Hundersmarck shared their educational-values hypothesis, contending that “the best teachers hold a particular set of values about education—typical examples include a commitment to helping all kinds of children learn, valuing diversity and caring, and espousing patience and persistence” (as cited in Metzgar & Wu, 2008, p. 921). Haberman (2005b) concurs, suggesting that there are critical attributes of teachers who are successful in urban schools, and that these teachers exhibit a unique set of beliefs and characteristics, including persistence, application of theory, approach to at-risk students, and fallibility.

Many scholars agree that the recruitment and selection of a more diverse teaching force is needed as we embrace the changing demographics of our nation's schools. While not all scholars agree that teachers of color are better able to work with diverse student populations, some research indicates that teachers of color are frequently better equipped to work effectively with diverse student populations (Eubanks & Weaver, 1999; Sleeter, 2001; Villegas & Lucas, 2002). Many

scholars also contend that teachers of color are better equipped to build positive relationships and connect with students of color (Gay, 2010; Ladson-Billings, 2009). Haberman (2005a) supports the call for more teachers of color, but also contends that teachers from urban backgrounds and those with experiences with diversity (including work with diverse student populations) should be recruited for teaching in urban schools. Other scholars contend that race is not a predictor of teacher quality, and that white teachers can build relationships with and effectively teach students from diverse populations (Sleeter & Thao, 2007).

Supporting Haberman's assertions, Pohan (1996) found that teacher education students "who bring strong biases and negative stereotypes about diverse groups will be less likely to develop the types of professional beliefs and behaviors most consistent with multicultural sensitivity and responsiveness" (p. 202). In addition, Garmon (2004) suggested that the character traits of preservice teachers also impact their potential for the development of multicultural awareness and sensitivity. Haberman and Post (1998) contended that "to perform the sophisticated expectations of multicultural teaching, selecting those predisposed to do it is a necessary precondition. Training, while vital, is only of value to teacher candidates whose ideology and predispositions reflect those of outstanding, practicing teachers" (p. 96). Given the current emphasis on teacher education accountability, the introduction of high CAEP standards, and the understanding that beliefs and values impact the potential for preservice teachers' success, EPPs must give critical thought to the process by which they admit students into their programs, and determine what policies should be enacted that will result in candidates who possess the necessary skills to effectively teach diverse student populations.

Non-Academic Skills

While teacher selection has been a focus of the literature for decades, the teacher education community has not yet settled on policies regarding the best way to select teachers or teacher candidates. However, the literature calls for a focus on non-academic skills in addition to high academic standards. With these calls come a variety of suggested approaches for measuring the non-academic skills of prospective teachers. Strong and Hindman (2006) developed a Teacher Quality Index (TQI) to assist school districts in identifying effective teachers. The TQI is an interview protocol designed to assess skills that cannot be "obtained from the employment application (the initial means of evaluation, the focus of which should be to determine if the applicant has the minimum qualifications for the position)" (Strong & Hindman, 2006, p. 5). The protocol attempts to assess pedagogical skills as well as personal characteristics, such as caring, motivation, enthusiasm, dedication to teaching, and reflective practice. Many educational organizations also attempt to measure candidate's non-academic abilities. Teach for America's (2016) selection process claims to assess skills such as critical thinking, leadership ability, perseverance, interpersonal skills, and respect for diversity. The New Teacher Project (TNTP, 2016) selection criteria include assessing for critical thinking, commitment to student achievement, professional interactions, and constant learning.

CAEP (2016a) also supports using non-academic criteria in selecting teacher candidates. Within the rationale for Standard 3, CAEP (2016a) states, "There is strong support from the professional community that qualities outside of academic ability are associated with teacher effectiveness," and that "Research has not empirically established a particular set of non-academic qualities that teachers should possess," yet, "The CAEP Commission recognizes the ongoing development of this knowledge base and recommends that CAEP revise criteria as evidence emerges" (p. 2). Some commercial instruments claim to measure the non-academic skills of prospective teachers. Table 1 lists instruments Ebmeier, Dillon, and Ng (n. d.) identified as "the most common commercial instruments" (p. 1).

Table 1
Common Commercial Screening Instruments

Instrument	Organization	Website
Insight	Kenexa	www.kenexa.com
Interactive Computer Interview System (ICIS)	American Association of Personal Administrators	www.aaspa.org
Star Teacher Interview	Haberman Foundation	www.habermanfoundation.org
Teacher Style Profile	Ventures for Excellence	www.venturesforexcellence.com
TeacherInsight	Gallup	www.gallup.com/consulting/education/22093/TeacherInsight.aspx

Source: Ebmeier, Dillon, and Ng, (n.d.)

After reviewing these instruments, Ebmeier et al. (n.d.) recommended careful consideration in selecting instruments consistent with the need of the organization and those that have affordable and high quality training components. In addition, the authors warned that “very few of the commonly available instruments have been published in peer reviewed journals and replication studies by external independent reviewers are lacking” (Ebmeier et al., p. 1). Therefore, it is vital that EPPs engage in critical examination of the use of instruments to determine their validity and utility as selection measures meeting the need for assessing the non-academic characteristics of prospective candidates. This need becomes even more vital when selecting candidates to teach in diverse, urban settings.

Context

Implemented in 2005 as a response to national and local calls for reform in urban teacher preparation, the urban teacher education program involved in the present study is a four-year undergraduate teacher preparation program in a large Midwestern city. The program was designed to expose candidates to the challenges and opportunities of urban education via the provision of a rigorous curriculum and experiences within urban schools. The mission of the program is to prepare exemplary teachers for urban schools by applying research in urban teacher preparation to curricular innovation, connecting theory to practice, and increasing clinical experiences in urban schools and communities (Waddell, 2015; Waddell & Ukpokodu, 2012).

Program Emphasis

To fulfill this mission, the four-year urban teacher preparation curriculum focuses on teaching for social justice and multiculturalism through a field-based model (Waddell, 2015; Waddell & Ukpokodu, 2012). Candidates are exposed to urban schools, teachers, students, and communities beginning the second week of their freshman year. Throughout the program, candidates are provided opportunities to explore their own culture, the cultures of urban communities, and the cultures of their students, making the exposure to and understanding of diversity central to the curriculum (Waddell, 2015; Waddell & Ukpokodu, 2012). The attention to urban communities includes learning about the history of urban schools within the United States and the districts within which the candidates will be employed. Candidates are encouraged to view the school system through the lens of social justice, creating an accurate understanding of the inequities extant in U.S. educational systems. A unique aspect of the curriculum is a summer course in which candidates are fully immersed in the urban community. Viewing the community through the lens of urban youth, families and community agencies help candidates take a strengths-based perspective. Through this

experience, they begin to discover the realities and assets of urban communities, the need for relationships with community stakeholders, and the inequities plaguing urban communities as a whole (Waddell, 2011, 2013).

Field Experience

Darling-Hammond (1997) identified several qualities of EPPs whose graduates were successful in teaching diverse learners effectively. She asserted that such programs had extended clinical experiences with “strong relationships, common knowledge, and shared beliefs among school- and university-based faculty,” and were “taught in the context of practice” (Darling-Hammond, 1997, p. 30). Recommendations for programs aiming to prepare teachers for urban schools included: (a) extended and carefully designed clinical components, (b) opportunities to work with diverse learners, and (c) fieldwork closely supervised and supported by clinical educators and mentors (Darling-Hammond, 1997). The National Council for Accreditation of Teacher Education (NCATE, 2010) echoed these recommendations and called on EPPs to partner with districts and schools, employ rigorous selection processes of candidates, and provide opportunities for candidates to work in hard-to staff schools. In an effort to fully prepare candidates for their futures as urban school teachers, the program involved in this study responded to the calls from the literature, and was designed to mimic the experiences of practicing teachers, providing candidates with extensive, hands-on knowledge of the realities and challenges of teaching in urban schools.

The program includes field experiences each semester of the four-year undergraduate program. During the first semester of the freshman year, the field experience is observation and exposure to the diversity of schools. However, by the sophomore year, the candidates are working as pre-service teachers one full day per week, serving as teacher’s aides in the classroom. During the summer prior to the final year, candidates work as interns in community agencies, gaining exposure to the realities of urban communities and the experiences of their students. In addition, the program does not employ typical student teaching; instead, the candidates are regarded as co-teachers during the final year of the program. The candidates are involved in a year-long co-teaching internship in which they begin the school year when the school district teachers return in August, and they follow the school district calendar for the academic year. During the year-long internship, the candidates work in the school three to five days per week in the fall semester, and full-time during the spring semester. In addition, much of the coursework is taught in classrooms in urban schools, and candidates are evaluated on their ability to apply coursework and work effectively with students. Therefore, the program provides candidates with hands-on experience as teachers in urban schools. One principal of a partner school stated that she hires graduates of the program, because “when they begin their first job, they already have teaching experience, it is as if I am hiring second-year teachers instead of first year teachers” (T. Degraff, personal communication, May 6, 2016).

Candidate recruitment and selection. Aware of the increased rigor of the program and the way it mimics the realities of urban teaching, administrators in the urban education program thought critically about the selection of candidates for the program. There was a need to select candidates who did not just have the propensity to be successful in a university teacher preparation program, but also with the propensity to be successful as an urban school teacher during and after the program. Therefore, the program approached selection and admission to the program in a manner consistent with the recommendations cited in the literature regarding the selection of urban teachers (Darling-Hammond & Baratz-Snowden, 2007; Haberman, 2005b; Ladson-Billings, 1995; Ryan & Alcock, 2002; Tredway, 1999; Weiner, 2000). The urban teacher education program targeted recruitment efforts at candidates of color, candidates with urban school experiences, candidates with a professed desire to teach in urban communities, candidates living in urban communities, and those

with experiences working with diverse populations of children (Haberman, 2005a; Waddell, 2015; Waddell & Ukpokodu, 2012).

The Selection Instrument

In addition to screening candidates for academic skills, the program also selected a commercially available instrument for use in measuring candidate's non-academic skills. The administrators felt that since the program was designed as a field-based program, in which success in the program was only possible if candidates were successful working in urban schools and classrooms, it would be wise to select candidates based on their potential for working in urban schools. The Haberman *Star Teacher Selection Interview* was selected because it was designed specifically to identify teachers who can successfully work with children in urban schools and/or children of poverty (Haberman, 2005b). The interview was designed to predict Star (successful) teachers for urban schools through addressing seven beliefs or characteristics. The intent of the Star Teacher Selection Interview is to predict a candidate's potential for success in urban schools by addressing seven primary functions: (a) persistence; (b) diplomatic response to authority; (c) application of theory and generalizations; (d) approach to at-risk students; (e) personal/professional orientation; (f) resilience in the face of burnout; and (g) fallibility. The scored interview attributes a higher score to those candidates whose response displays the respective function in a manner consistent with successful urban teachers. According to Haberman (2005b), this selection instrument distinguishes Star teachers from those destined to quit or fail in urban schools. The interview data reported by Haberman (personal communication, July 8, 2009) suggested that the interview is able to predict who succeeds and persists, as well as who quits or fails as a teacher in an urban school. According to Haberman (2005b), 90% of candidates who fail the interview are also those who, if they become teachers, will either leave the profession or be ineffective in the classroom, while 95% of candidates who pass the interview become teachers who are effective and persist in employment in urban schools. Haberman (2005b) asserts that "all those who pass the interview have the predispositions to succeed as urban teachers serving diverse students in poverty" (p. 91).

Purpose

Given the calls from the literature regarding selectivity in teacher education, accountability, and programmatic reforms aimed at better preparing teachers for urban schools, and the success rate of graduates from the EPP involved in this study, the purpose of the study was to explore the validity of an existing commercial teacher selection instrument as a selection instrument for EPPs. The *Standards for Educational and Psychological Testing*, written jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), discusses five sources of evidence for the validity of measurements, including evidence based on: (a) test content, (b) response processes, (c) internal structure, (d) relations to other variables, and (e) the consequences of test use (AERA, APA, & NCME, 2014). The present study focused on evidence regarding internal structure, specifically in the form of exploratory factor analysis, (see Standard 1.13, AERA et al., 2014, pp. 26-27), and evidence regarding relationships with criteria in the form of logistic regression (see Standards 1.17-1.19, AERA et al., p. 28). The *Standards* note that the reliability of a measurement impacts its validity (AERA et al., pp. 34-35), and thus, the present study sought to obtain evidence of internal consistency reliability in the form of Cronbach's alpha (see Standard 2.6, AERA et al., p. 44) and inter-rater reliability in the form of Krippendorff's alpha (see Standard 2.7, AERA et al., p. 44).

The current policy debate regarding CAEP standards and selectivity in teacher education warrant assessment and evaluation of the evidence for validity of current and proposed selection procedures. Based on the literature reviewed, we hypothesize that: (a) a commercial teacher selection instrument designed for use in selecting teachers for urban schools will predict the success of students within a field-based urban teacher preparation program; (b) ACT score will be a predictor of success in the program; (c) minority status of the student will not be a significant predictor of scores on the interview or success in the program; (d) students with an urban background will score better on the commercial teacher selection instrument; and (e) students with urban backgrounds will be more likely to experience success in the urban teacher preparation program. Testing these hypotheses will inform the policy debates regarding EPP admissions and selectivity.

Methodology

Participants

Data were available from 109 students over a six-year period (2005-2011) and had been collected as a component of their admission to a teacher education program. The sample consisted of 21 males and 88 females, and the ethnicity of the participants was 48% African American, 45% Caucasian, 6% Hispanic, and 1% Asian. Of the total sample used in the study, 73 of the participants completed the urban teacher education program, and 36 participants left the program prior to completion for a variety of reasons, including voluntary separation due to personal problems outside of the classroom (14 participants, 12.8%), involuntary separation due to poor academic performance (6 participants, 5.5%), involuntary separation due to failure to consistently meet the professional standards of the program (6 participants, 5.5%), or some combination of these reasons (10 participants, 9.2%). Participants were coded as Graduates or Leavers for data analysis, and further demographic decompositions can be seen in Table 2. ACT score was available for 80 students, and ranged from 14 to 28 ($M = 19.56$, $SD = 3.54$).

Table 2
Demographic Variables by Reason for Leaving the Program (N = 109)

	Graduated		Other Reason		Total	
	<i>f</i>	%	<i>F</i>	%	<i>f</i>	%
Gender						
Male	9	42.86	12	57.14	21	100.00
Female	64	72.73	24	27.27	88	100.00
Race/Ethnicity						
White	40	81.63	9	18.37	49	100.00
Black	30	57.69	22	42.31	52	100.00
Hispanic	2	28.57	5	71.43	7	100.00
Asian	1	100.00	0	0.00	1	100.00
Previous School Background						
Urban	44	58.67	31	41.33	75	100.00
Suburban	18	94.74	1	5.26	19	100.00
Rural	5	71.43	2	28.57	7	100.00
Private	4	80.00	1	20.00	5	100.00
Unknown	2	66.67	1	33.33	3	100.00
Class at Entry to Program						
Freshman	41	54.67	34	45.33	75	100.00
Sophomore	17	89.47	2	10.53	19	100.00
Junior	15	100.00	0	0.00	15	100.00
Year of Entry to Program						
2005	7	63.64	4	36.36	11	100.00
2006	10	71.43	4	28.57	14	100.00
2007	9	56.25	7	43.75	16	100.00
2008	13	72.22	5	27.78	18	100.00
2009	13	59.09	9	40.91	22	100.00
2010	19	73.08	7	26.92	26	100.00
2011	2	100.00	0	0.00	2	100.00

Instrumentation

The *Star Teacher Selection Interview* (Haberman Educational Foundation, Inc., 1994) is an interview instrument designed to assess the potential of a candidate for success, effectiveness, and persistence in an urban school setting.

The Haberman Teacher Selection Interview assesses the seven characteristics/beliefs of effective urban teachers through separate two-part questions. The instrument consists of a series of 15 open-ended stems and responses scored on a 0-3 point scale with 0 equated with failure. Scores are totaled for an overall score, and are categorized according to potential for successful teaching in urban schools. Table 3 shows which questions and subquestions screen for which characteristics or belief systems.

Table 3
Haberman Teacher Selection Interview Functions and Subquestions

Function	SubQuestion	Characteristic/Belief System
Persistence	Persistence A	Application of Persistence
	Persistence B	Understanding of the Need for Persistence
Diplomatic Response to Authority	Response to Authority A	Diplomatic Advocacy for Learning
	Response to Authority B	Diplomatic Response to Authority
Application of Theory and Generalizations	Application A	Understanding of Theory
	Application B	Application of Theory
	Application C	Transference of Practice to Theory
Approach to At-Risk Students	At-Risk A	Understanding of At-Risk Challenges
	At-Risk B	Approach to At-Risk Students
Personal vs. Professional Orientation	Orientation A	Application of Professional Orientation to Teaching
	Orientation B	Understanding of Professional Orientation
Resilience in the face of Burnout	Burnout A	Understanding of Teacher Burnout
	Burnout B	Displays Resilience
Fallibility	Fallibility A	Recognizes Own Fallibility
	Fallibility B	Displays Humility Regarding Fallibility

According to Star Teacher Selection Interview Training Manual (Haberman Educational Foundation, 1994), a total score of 40 or higher earns the candidate the title of Star Teacher, while a score of 0 on any item indicates failure of the entire instrument. Haberman (2005b) reports inter-rater reliability, test-retest reliability, and predictive validity evidences for the instrument. Haberman (2011) stated that, “in terms of the reliability of interview teams, when using trained teams, the interviewers become reliable after six joint interviews; that is, each will score an interview within four points (out of a possible 45 perfect total score) in 80% of the cases. After six joint interviews, the interviewers will pass (or fail) the same applicants in 95% of the cases” (para. 2).

As indicated previously, a criticism of the Haberman instrument is the fact that there have not been external empirical studies published to support Haberman’s claims (Ebmeier, Dillon, & Ng, n.d.; Metzgar & Wu, 2008). Because CAEP standards dictate the use of reliable and valid selection measures and the literature calls for selecting candidates with academic and non-academic skills, the administrators of the program were interested in assessing the evidence for validity of the instrument scores within the teacher candidate population. Therefore, the use of the instrument for this study attempts to respond to the calls in the literature for selectivity in teacher education, and to provide empirical data regarding the validity of the instrument for use in teacher education.

Procedure

Data for this study were collected as a component of the admissions process into the urban education program. Students were recruited for the program through college recruitment fairs and campus visits, recruiting visits to partner schools, the university website, partner district and community partnership recruiting venues, advertising on billboards and radio (in some years) and word of mouth. The admissions process included an application to the program, a review of transcripts and academic test scores, a personal essay, letters of recommendation and participation in two personal interviews (Waddell, 2015; Waddell & Ukpokodu, 2012). The first interview was one in which applicants responded to questions regarding their interest in teaching in urban schools as well as questions about their strengths, challenges and experiences; this interview was administered by faculty members in groups of two to three. The second interview was the Haberman Star Teacher Selection Interview. This interview was administered by faculty members in groups of two, and each faculty member participating in the interview had been trained by the Haberman Educational Foundation, and thereafter received a certificate of reliability from the Foundation.

Scores were recorded to be used as (a) part of a longitudinal evaluation of the program and (b) an indicator of the candidate's potential for successful teaching in urban schools. The scores themselves were not used for admissions decisions. The evaluation plan for the program included administering the interview to all candidates again as an exit interview at the time of graduation to determine if changes in the scores occur after participating in the four-year program. The program has used the interview for 13 years (2005-2017), with 9 of the 13 cohorts completing the program. In examining available data from this period, trends have emerged regarding the responses to particular questions, showing potential for predicting success and/or failure in the program itself. Therefore, the purpose of this study was to explore the viability and validity of the instrument as a predictor for success in an urban teacher education program.

Results

Preliminary Analyses

All analyses were conducted using SPSS v23. One case had what appeared to be a data entry error for Persistence B (an impossible score of 5), and expectation maximization was used to impute the missing score. A chi-square test indicated that students who left the program without graduating were disproportionately Black (61% versus 25% White, 14% Hispanic, and 0% Asian; $\chi^2[3] = 17.95$, $p < .01$; Cramer's $V = .33$, a moderate effect). A significant difference in gender was also found ($\chi^2[1] = 6.84$, $p < .01$), with 57% of males leaving and 27% of females. No differences were found among cohorts ($\chi^2[6] = 3.28$, $p = .77$), or secondary school types ($\chi^2[4] = 9.41$, $p = .052$). A significant moderate difference in mean ACT score was found between those who remained in the program ($M = 20.31$, $SD = 3.12$) and those who left ($M = 18.39$, $SD = 3.88$, $t[78] = 2.44$, $p = .02$, $.95CI = [0.35, 3.49]$). Those who left the program scored about half a standard deviation lower than those who remained (Cohen's $d = 0.56$).

Table 4 summarizes correlations among the Haberman subquestions, ACT score, program status (whether or not someone left the program), and dichotomized versions of age at entry (0 = less than 20 years old, 1 = 20 years old or older), ethnicity (0 = White, 1 = minority), and high school attended (0 = Other, 1 = Urban public). The demographic variables correlated significantly to one another with few exceptions, and to program status. Each also correlated significantly with at least one of the Haberman subquestions. In particular, Burnout A and Burnout B had moderate to strong correlations with each of the demographic variables and program status. The Haberman

subquestions on the whole lacked significant relationships with one another. With few exceptions, the largest correlations occurred between subquestions belonging to the same topic (or question). For example, At Risk A and At Risk B had a strong positive correlation of .52 ($p < .01$). The lack of significant zero-order correlations may have been due to a lack of reliability, which was assessed next. Associations among the questions may also be stronger once shared variance among them is accounted for, and this was later assessed with exploratory factor analysis.

Structural Validity of the Haberman

Inter-rater reliability. A total of eight raters were used to score Haberman interview responses on the 15 questions and subquestions, and 92 student responses were scored by at least two raters (most were scored by two, but some by more). Agreement among the raters was assessed using Krippendorff's alpha, α_K , through a SPSS macro employing 1000 bootstrapping samples (Hayes & Krippendorff, 2007). Reliability coefficients ranged from .79 to .96, with the majority falling above .90 ($M = 0.91$, $SD = 0.04$). Table 5 summarizes the coefficients for each of the questions. Overall, good evidence of inter-rater reliability was found.

Exploratory factor analysis. An exploratory factor analysis (EFA) using principle axis factoring with Promax rotation was conducted on the Haberman scores. In accord with the seven-question structure of the Haberman, the Kaiser criterion (i.e., the number of eigenvalues ≥ 1.00 ; the seventh eigenvalue in this analysis = 0.99) and scree plot both supported the extraction of seven factors, which explained 45% of the variance in the question scores. Just 7% of the correlation residuals were greater than $|\ .05 |$. Communalities ranged from small (e.g., .07) to large (e.g., .72). Table 6 summarizes the pattern and structure coefficients and communalities. Although seven factors were extracted, they did not align perfectly with the seven questions of the Haberman. For example, Factor 4 seems to be the main driver behind responses to just two out of the three Application subquestions, Factor 5 accounts for just one of the two Response to Authority subquestions, and Factor 7 to just one of the two Persistence subquestions. Factor 6 seems to underlie response to both of the remaining Persistence and Application subquestions. However, Factors 1-3 account for clearly delineated groups of Fallibility, At Risk, and Burnout subquestions, respectively, as intended by the Haberman developers. None of the factors explained responses to Response to Authority B or either of the Orientation subquestions very well; none had pattern or structure coefficients above .27, nor a communality above .18. These results seemed to echo the pattern of zero-order correlations discussed in Preliminary Analyses and displayed in Table 4.

Table 4

Correlations, Means, and Standard Deviations for Haberman Subquestions, Age, Minority Status, Urban Background, Program Status, and ACT Score ($N = 109$).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Persistence A	—																		
2. Persistence B	.21*	—																	
3. Authority A	-.01	.13	—																
4. Authority B	-.04	.16	-.05	—															
5. Application A	.07	.09	.26†	.07	—														
6. Application B	.04	.21*	.16	.07	.42‡	—													
7. Application C	.20*	.09	.13	.09	.02	.20*	—												
8. At risk A	-.06	.11	.00	.17	.02	.01	.17	—											
9. At risk B	-.06	.05	.09	.13	.11	.13	.12	.52‡	—										
10. Orientation A	.14	.00	-.10	.11	.06	.09	.09	.06	.07	—									
11. Orientation B	.06	.03	.12	.06	-.05	-.02	.08	.10	-.01	.10	—								
12. Burnout A	-.01	.09	.23*	.17	.08	.06	.10	.31†	.24†	.15	.09	—							
13. Burnout B	.11	.13	.01	.25†	.12	.10	.13	.39‡	.41‡	.14	.01	.49‡	—						
14. Fallibility A	-.09	.16	.30†	.16	.23*	.18	.04	-.07	.11	-.09	.12	.08	-.02	—					
15. Fallibility B	.07	.11	.19	.15	.03	.01	.29†	.03	.25†	.01	.14	.08	.06	.51‡	—				
16. Age at entry ^a	.04	.15	.11	.09	.17	.05	.03	.43‡	.26†	-.11	.09	.34‡	.40‡	.01	.10	—			
17. Minority ^b	-.01	-.07	-.18	-.16	-.05	-.17	.02	-.18	-.24*	-.03	-.15	-.25†	-.43‡	-.07	-.16	-.13	—		
18. Urban ^c	-.11	-.11	-.05	-.07	-.19	-.24*	-.07	-.17	-.28†	-.13	-.04	-.24*	-.42‡	-.04	-.23*	-.13	.51‡	—	
19. Status ^d	-.14	-.11	-.18	-.30‡	-.19*	-.12	-.06	-.17	-.18	-.06	.15	-.37‡	-.60‡	-.09	.06	-.19*	.28†	.26†	—
20. ACT	.11	.14	.10	.12	.20	.19	-.06	-.06	.21	.19	-.03	.28*	.25*	.07	.17	-.01	-.55‡	-.39‡	-.27*
<i>M</i>	2.39	2.59	1.38	1.52	2.03	1.76	1.63	1.20	1.55	2.19	2.31	1.06	1.12	1.67	1.90	0.32	0.55	0.69	0.33
<i>SD</i>	0.58	0.49	0.86	0.88	0.58	0.58	0.59	0.88	0.68	0.60	0.72	0.88	0.79	0.73	0.63	0.47	0.50	0.47	0.47

Note. ACT: $n = 80$, $M = 19.56$, $SD = 3.54$. Authority = Response to authority.

a. 0 = Under 20 years, 1 = Over 20 years. b. 0 = White, 1 = Minority. c. 0 = Other high school, 1 = Urban public high school. d. 0 = Remained in program, 1 = Left program.

* $p < .05$. † $p < .01$. ‡ $p < .001$.

Table 5
Krippendorff's Alpha Inter-Rater Reliabilities for Haberman Subquestions (N = 92)

Subquestion	Est. α_K	.95CI		$H_0 \alpha_K^a$				Units	Observers	Pairs
		LB	UB	.90	.80	.70	.67			
Persistence A	.89	.83	.94	.61	.01	.00	.00	86	8	96
Persistence B	.91	.84	.97	.35	.00	.00	.00	84	8	94
Response to Authority A	.92	.87	.96	.18	.00	.00	.00	86	8	94
Response to Authority B	.95	.93	.97	.00	.00	.00	.00	86	8	96
Application A	.79	.70	.88	.99	.54	.02	.01	84	8	94
Application B	.87	.80	.92	.80	.03	.00	.00	86	8	96
Application C	.90	.84	.94	.54	.00	.00	.00	85	8	95
At Risk A	.92	.84	.97	.27	.00	.00	.00	87	8	97
At-Risk B	.88	.79	.93	.72	.04	.00	.00	87	8	97
Orientation A	.90	.86	.94	.50	.00	.00	.00	86	8	96
Orientation B	.89	.80	.95	.60	.02	.00	.00	85	8	97
Burnout A	.95	.92	.97	.00	.00	.00	.00	86	8	96
Burnout B	.93	.90	.96	.05	.00	.00	.00	86	8	98
Fallibility A	.96	.95	.98	.00	.00	.00	.00	86	8	96
Fallibility B	.94	.91	.97	.01	.00	.00	.00	86	8	96

Note. Krippendorff's alpha is symbolized with α_K .

^aProbabilities associated with the null hypothesis test that the population parameter of Krippendorff's alpha is a particular value.

Table 6
Factor Coefficients and Communalities for Haberman Subquestions (N = 109).

Subquestion	Factor							h^2
	1	2	3	4	5	6	7	
Persistence A	-.05(-.02)	-.18(-.05)	.03(.06)	.04(.06)	-.01(-.04)	.49(.48)	.15(.21)	.28
Persistence B	.01(.16)	.03(.09)	-.03(.16)	-.01(.22)	.04(.08)	.08(.17)	.76(.77)	.60
Response to Authority A	.03(.30)	.01(.04)	.07(.14)	.09(.29)	.77(.80)	.03(-.02)	.03(.13)	.67
Response to Authority B	.21(.21)	.04(.21)	.25(.31)	.04(.13)	-.22(-.13)	-.04(.05)	.11(.20)	.18
Application A	-.02(.15)	-.01(.06)	.03(.14)	.65(.65)	.09(.22)	.00(-.02)	-.06(.14)	.44
Application B	-.05(.13)	.04(.10)	-.07(.11)	.66(.65)	.01(.13)	.13(.12)	.05(.24)	.45
Application C	.14(.21)	.12(.23)	-.04(.12)	.06(.11)	.10(.11)	.48(.49)	-.03(.08)	.30
At Risk A	-.12(.01)	.68(.70)	.11(.43)	-.10(.00)	.03(-.03)	-.04(.13)	.08(.10)	.52
At Risk B	.09(.23)	.79(.77)	-.05(.35)	.10(.18)	.00(.03)	-.07(.09)	-.05(.03)	.62
Orientation A	.00(-.03)	-.10(.09)	.25(.23)	.12(.09)	-.18(-.16)	.25(.27)	-.11(-.01)	.16
Orientation B	.15(.16)	-.06(.04)	.13(.12)	-.12(-.05)	.09(.11)	.12(.13)	-.01(.03)	.07
Burnout A	.00(.10)	-.02(.36)	.78(.76)	-.08(.10)	.19(.20)	-.02(.09)	-.05(.10)	.61
Burnout B	-.05(.04)	.26(.54)	.51(.66)	.06(.17)	-.10(-.09)	.06(.22)	.03(.17)	.52
Fallibility A	.73(.76)	-.11(.01)	.05(.07)	.14(.34)	.04(.32)	-.24(-.22)	.07(.22)	.68
Fallibility B	.82(.79)	.07(.22)	-.05(.06)	-.14(.04)	.00(.19)	.23(.26)	-.05(.09)	.72

Note. Pattern coefficients are listed first, followed by structure coefficients in parentheses. Coefficients loading onto a particular factor are boldfaced. Communalities are represented by h^2 . Variance explained = 45%. Factor correlations were as follows: $r_{12} = .17$, $r_{13} = .10$, $r_{14} = .25$, $r_{15} = .30$, $r_{16} = .04$, $r_{17} = .19$, $r_{23} = .50$, $r_{24} = .11$, $r_{25} = .50$, $r_{26} = .22$, $r_{27} = .08$, $r_{34} = .20$, $r_{35} = .05$, $r_{36} = .17$, $r_{37} = .21$, $r_{45} = .22$, $r_{46} = -.01$, $r_{47} = .29$, $r_{56} = -.09$, $r_{57} = .07$, and $r_{67} = .13$.

Evidence for total score. Overall, the EFA results provide good evidence to support a multifactor structure for the Haberman, but that structure may not reflect its originally intended use. Using a total score would have been supported by the extraction of a single factor rather than seven. An EFA was performed on the factor correlation matrix to support the existence of a second-order factor, which would also support the use of a total score. However, two factors were extracted that accounted for a mere 29% of the variance in the factors, failing to support a hierarchical structure. In addition, Cronbach's alpha was computed for the total scale, and found to be .65, a poor value for internal consistency, and indicative that Haberman total scores may contain substantial (i.e., 35%) random error. Again, this aligns with the pattern of correlations discussed earlier (see Table 4).

Main Analysis

Initial independent t-tests were performed comparing each group, those who remained in the program (coded as 0) and those who left (coded as 1), on each of the Haberman subquestions. Significant differences were found for Response to Authority B ($M = 0.55$, $SE = 0.11$, $t[107] = 3.21$, $p < .01$, $.95CI = [0.21, 0.89]$), Application A ($M = 0.24$, $SE = 0.12$, $t[107] = 2.02$, $p = .046$, $.95CI = [0.004, 0.47]$), Burnout A ($M = 0.68$, $SE = 0.15$, $t[95.90] = 4.62$, $p < .01$, $.95CI = [0.35, 0.98]$), and Burnout B ($M = 1.01$, $SE = 0.11$, $t[96.49] = 8.90$, $p < .01$, $.95CI = [0.79, 1.24]$). These results mirrored the significant correlations between program status and the Haberman subquestions discussed on Preliminary Analyses and displayed in Table 4, and also supported the exploration of the aggregate power of these subquestions to predict whether an applicant remained in or left the program.

Logistic regression. To that end, we entered all 15 subquestions as independent variables in a logistic regression model predicting whether a student remained or left. Two influential multivariate outliers were identified and removed, because model fit substantially improved after their removal ($AIC = 100.75$ for the model with outliers, and 77.23 for the model without them). To account for possible dependency in the data, dummy-codes for cohort were entered in the initial block, but there was no significant change in model fit— $\Delta \chi^2(5) = 8.96$, $p = .11$ —and cohort was not modeled in subsequent analyses. This model greatly improved prediction over the null model (Nagelkerke $R^2 = .79$), and correctly predicted whether a student would leave or stay in 89% of the cases (both 89% of those who remained, and 89% of those who left). Significant predictors in the model included Response to Authority B ($OR = 0.23$), and Burnout B ($OR = 0.006$; see Table 5 for model parameter estimates).

Because the demographic variables of age at entry, ethnicity, and high school attended were correlated with both the dependent variable of program status and at least one of the predictors, it was important to control for them, and thus they were entered into a second model block. Their addition did not significantly change the fit of the model ($\Delta \chi^2[3] = 7.13$, $p = .07$) nor add much to its predictive power (Nagelkerke $R^2 = .83$, correct classification rate = 91%). However, controlling for those variables revealed several additional Haberman subquestions as having a significant association with eventual program status, such as Burnout A ($OR = 0.10$), Fallibility A ($OR = 0.05$), and Fallibility B ($OR = 20.73$; see Table 7). Response to Authority B remained significant, albeit with a larger effect size ($OR = 0.11$), as did Burnout B ($OR = 0.003$). When all other variables were controlled, the odds of leaving the program decreased relative to the odds of remaining by a factor of 9 for every standard deviation increase in Response to Authority B, decreased by a factor of 10 for every standard deviation increase in Burnout A, a factor of 333 for every standard deviation increase in Burnout B, and a factor of 18 for every standard deviation increase in Fallibility A. Scoring higher on each of these Haberman subquestions was strongly associated with remaining in the program. On the other hand, the odds of leaving the program increased relative to the odds of remaining by a factor of over 20 for every standard deviation increase in Fallibility B. In other

words, scoring higher on this subquestion was strongly associated with leaving the program, an unexpected result.

Table 7

Logistic Regression Model of Haberman Star Interview Question Responses on Leaving^a an Urban Teacher Education Program Before Graduation (N = 107)

Variable	Model 1		Model 2		Model 3		Model 4 ^b	
	B(SE)	OR	B(SE)	OR	B(SE)	OR	B(SE)	OR
Intercept	4.75(3.61)	115.01	-2.31(4.91)	0.10	2.20(1.64)	9.04	6.31(3.67)	552.42
Persistence A	0.38(0.88)	1.46	0.21(1.03)	1.23				
Persistence B	-0.47(1.08)	0.62	-0.37(1.52)	0.69				
Authority A	-0.94(0.58)	0.39	-1.24(0.75)	0.29				
Authority B	-1.46(0.57)*	0.23	-2.19(0.86)*	0.11	-1.00(0.41)*	0.37	-0.97(0.44)*	0.38
Application A	-1.03(0.82)	0.36	-0.74(0.98)	0.48				
Application B	-0.31(0.87)	0.74	0.11(1.08)	1.12				
Application C	-0.14(0.77)	0.87	-0.87(0.99)	0.42				
At risk A	-0.22(0.64)	0.80	-0.48(0.82)	0.62				
At risk B	1.53(0.90)	4.60	2.47(1.32)	11.81				
Orientation A	0.28(0.64)	1.32	0.90(0.79)	2.46				
Orientation B	1.34(0.75)	3.81	2.39(1.29)	10.88				
Burnout A	-1.87(1.01)	0.15	-2.35(1.17)	0.10	-1.26(0.59)*	0.28	-0.86(0.64)	0.42
Burnout B	-5.08(1.51)†	0.01	-5.89(1.98)‡	0.00	-3.28(0.80)‡	0.04	-3.58(1.02)‡	0.03
Fallibility A	-1.45(0.75)	0.24	-2.92(1.22)*	0.05	-1.38(0.57)*	0.25	-1.56(0.66)*	0.21
Fallibility B	1.46(0.81)	4.32	3.03(1.30)*	20.73	1.66(0.62)†	5.28	1.59(0.63)*	4.91
Age at entry ^c			-0.92(1.39)	0.40	0.28(0.90)	1.33	0.19(1.30)	1.21
Minority ^d			1.70(1.55)	5.45	0.20(0.80)	1.22	-0.69(1.00)	0.50
Urban ^e			3.39(2.05)	29.72	1.33(1.02)	3.77	0.72(1.11)	2.06
ACT							-0.14(0.13)	0.87
$\chi^2(df)$	88.92(15)‡		96.04(18)‡		76.02(8)‡		55.21(9)‡	
Nagelkerke R ²	.79		.83		.71		.68	

Note. The fit of Model 2 was not significantly different than Model 1: $\Delta\chi^2(3) = 7.13, p = .07$. The fit of Model 3 was significantly different than Model 2: $\Delta\chi^2(10) = 20.02, p = .03$.

a. 0 = Remained in program, 1 = Left program. b. N = 79. c. 0 = Under 20 years, 1 = Over 20 years. d. 0 = White, 1 = Minority. e. 0 = Other high school, 1 = Urban public high school.

* $p < .05$. † $p < .01$. ‡ $p < .001$.

One explanation for the high predictive power of the model may be the inclusion of so many nonsignificant predictors, or even simply having so many parameters, 19, relative to the sample size. Therefore, we trimmed the model to the five Haberman subquestions that had a significant coefficients and the three demographic covariates. The simpler model with 9 parameters fit significantly better than the more complex 19 parameter model ($\Delta\chi^2[10] = 20.02, p = .03$), although some predictive power was lost (Nagelkerke R² = .71, correct classification rate = 88%). Each Haberman subquestion remained significantly associated with program status in the same directions as before, although with weaker effect sizes (i.e., OR's closer to 1.00): Response to

Authority B (OR = 0.37) Burnout A (OR = 0.28), Burnout B (OR = 0.04), Fallibility A (OR = 0.25), and Fallibility B (OR = 5.28; see Table 7).

ACT score. We also fit the simpler model after adding ACT score, but could only use data from the 80 students for whom we had ACT scores. One of the two outliers identified previously was also identified as an influential outlier in the revised model and sample, and was removed. The resulting model had good predictive power (Nagelkerke $R^2 = .68$, correct classification rate = 86%), but ACT score was not a significant predictor ($B = -0.14$, $SE = 0.13$, $p = .26$, OR = 0.87). Even after controlling for ACT score, the same Haberman subquestions remained significant (except for Burnout A), with associations in the same directions, albeit with smaller effect sizes (see Table 7).

Haberman total score. Although no evidence was found to support the structural validity of a total score, we modeled the use of such a score for predicting program status because that is clearly the intention of the test developers. A model was fit using just a total Haberman score as a predictor (i.e., a summation of all subquestion scores; $M = 26.33$, $SD = 4.42$, $n = 109$), but it did not predict the data as well (Nagelkerke $R^2 = .30$, correct classification rate = 73%), although the total score was a significant predictor ($B = -0.30$, $SE = 0.07$, $p < .01$, OR = 0.74). When the covariates of age at entry, ethnicity, and high school attended were added to the model, the total score remained significant, but the correct classification rate decreased slightly to 71%.

Because the recommendations for the Haberman state that a score of zero on any one subquestion should result in a score of zero for the entire instrument, that version of the total score ($M = 15.09$, $SD = 14.53$, $n = 109$, 51 scores of 0 [47%]) was also modeled. The zero-scored Haberman had a statistically significant association with program status ($B = -0.06$, $SE = 0.02$, $p < .01$), but it was weak (OR = 0.94). Each one point increase in the score was associated with a reduction in the odds of leaving the program by just 6%. In addition, the predictive power of the model was poor: the correct classification rate increased from 67% for the baseline model to just 68%.

Discussion

Study results indicated that Haberman subquestion scores had fair to good inter-rater reliability, but there was poor evidence for the structural validity of a single total score. Internal consistency was less than adequate, and factor analysis indicated a multi-factor structure rather than a single factor structure. Likewise, predictive validity was poor for the total score, whether computed with simple summation or with the recommended scoring method. Logistic regression models improved classification rates over the baseline model (i.e., assuming all students would remain in the program) by only one to three percentage points. However, subquestion scores greatly improved classification rates (from 67% for the baseline model to 86-91%), even after controlling for minority status, type of high school attended, age at program entry, or ACT score. This finding is significant, in that it indicates that non-academic factors are a necessary complement to academic admissions standards for EPPs.

One limitation of the study is the small number of participants relative to the number of parameters in the analyses. However, those subquestions significantly associated with program status—Response to Authority B, Burnout A, Burnout B, Fallibility A, and Fallibility B—seemed to have robust effects: their regression coefficients remained significant through a variety of models, as did their effect sizes. Each one point increase in subquestion score (except Fallibility B) was associated with a several-fold increase in the odds of remaining in the program. However, it should

be noted that each subquestion score ranged from zero to three points, so there was a limited amount of movement in the scale.

Fallibility B was the only significantly-related subquestion to have a positive association with leaving the program, a surprising observation given that all other subquestions and demographic variables of interest were controlled (the association without those controls is weak and nonsignificant, $r = .06$). This result warrants further investigation, because Haberman (2005a) contended that higher scores for any of the questions indicate higher propensity for success in urban schools. However, because the present study examined the use of the instrument with a different population than Haberman's earlier studies, future studies with larger sample sizes should further examine Fallibility B and its association with attrition in teacher education programs. In addition, Haberman (2005a; personal communication, July 8, 2009) asserted that life experiences and maturity are predictors of high scores and success on the instrument. While positive association between Fallibility and attrition was obtained after controlling for age, age itself may not accurately represent an individual's maturity or the qualities of an individual's life experiences. Again, future studies should revisit this association and the measurement of maturity and life experience.

The sample size also seemed to constrain results in terms of power, because several variables had large effect sizes and yet remained statistically unrelated to program status. For example, Orientation B had an odds ratio of 3.81 (10.88 when demographic variables were controlled), but failed to reach significance. At Risk B had similarly large effect sizes without reaching significance. Future studies with larger sample sizes might find these and other variables to be associated with program status. Exacerbating the issue for minority status, type of high school attended, age at program entry was the dichotomization of these variables, which lost valuable information. However, cell sizes were too small for valid analysis using the original versions of these variables, a challenge that may be overcome with larger sample sizes.

Although the results seem to indicate that valid predictions about program success can be made with just the five subquestions, readers should note that responses to the questions and scoring of the questions are not independent. All are created within the overall context of the Haberman interview. Thus, Model 2 is recommended as the model to retain for future investigation. However, based on the assessment of the evidence for its structural validity, use of the total score cannot be recommended. In fact, due to the unexpected direction for the association of responses to Fallibility B with leaving the program, Haberman users are encouraged to take into account each subquestion score in a holistic fashion when using the instrument to support decisions about likely student success in urban education teacher preparation programs.

As noted previously, the Urban Teacher Education Program recruits candidates who have a stated desire to teach in urban schools. In addition, as part of the program, candidates receive financial support in the form of scholarships in exchange for a commitment to work in one of the program's partner school districts for a minimum of four years upon graduation. Therefore, the study is not transferable to all teacher education programs or those that recruit students professing to want to teach in rural or suburban communities. However, with the increasing diversity of communities across the United States, it is improbable that graduates of teacher preparation programs will not teach diverse populations. This limitation, then, does not pertain to the intent of programs but with the stated desires of applicants to the programs. Therefore, programs committed to preparing teachers for diverse populations in the United States may consider using the instrument in their programs. In addition, the success rates of program graduates coupled with the design of the program demonstrate that the instrument is relevant for selecting teachers for urban, field-based teacher preparation programs.

A large majority of participants in the study, both graduates and leavers, possessed some of the background characteristics Haberman (2005b) deemed predictive of individuals who will be successful teachers in urban schools. These participants (a) lived in or “were raised in a metropolitan area,” (b) “attended schools in a metropolitan area as a child or youth,” (c) were “African American, Latino, members of a minority group, or from a working class white family,” (d) had lived in poverty or could empathize with the challenges therein, or (e) had experience “working with children of diverse backgrounds” (Haberman, 2009, p. 82). Therefore, the sample is not generalizable to all teacher education programs.

Educational Implications

As previously stated, the *Standards for Educational and Psychological Testing* (AERA et al., 2014) specifically discuss five sources of validity evidence, of which two were explored in the present study, internal structure and relationships with criteria. Readers should note that the validity of the use of a measurement is itself a scientific theory (Furr & Bacharach, 2014), and therefore, the process of evaluating evidence for/against validity is ongoing. Users of the Haberman (e.g., scholars, policymakers, and education practitioners), therefore, should not regard the results of this study as proof of the instrument’s validity. Rather, they should regard themselves as contributors to the case for the validity of the use of its scores for an expanding array of applications. This can be accomplished by addressing those sources of validity that were unaddressed by this study, or by adding to the sources of evidence already obtained.

A case in point is the assessment of evidence for internal structure validity. Standard 1.13 (AERA et al., 2014, p. 26) states, “If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided.” As demonstrated by the EFA results and Cronbach’s alpha, there is a lack of structural validity evidence for the use of the Haberman as a single score. As noted above, the recommendation of the authors, therefore, is to approach any such application with caution, and that future test-users continue to gather evidence. Regarding relationships with criteria, however, the present study provides ample evidence. Standards 1.17-1.19 (AERA et al., 2014, p. 28) require test-users who justify the use of a test with observed associations with a criterion or criteria to report information about the criteria, how the criteria relate to the test, and how well other variables assist the test scores with predicting the criteria. The logistic regression results provide potential test-users with just such information. However, additional evidence is needed for increased generalizability.

Reliability has direct implications for validity, as well, and Standards 2.6 and 2.7 (AERA et al., 2014, p. 44) state that various reliability coefficients should not be assumed to be equivalent, and that evidence of inter-rater consistency should be provided when subjective scoring is used. Thus, present study provides two types of reliability, internal consistency and inter-rater reliability, and users of the Haberman can use the results with reasonable confidence to support their own uses of the instrument for the selection of teacher education candidates.

One possible criterion for selection is successful completion of the program, and the most accurate predictions of program success in the current study were obtained using the model that included all the Haberman subquestions and the covariates of age, minority status, and urban status (i.e., Model 2 in Table 6). Results indicated that if one assumed that all those admitted to the program would graduate (i.e., the baseline assumption), one would have been correct 67% of the time, but if one used the scoring procedure presented in this study (i.e., using each of the 15 Haberman items as predictors rather than a total score), one would have been correct 89% of the

time, an improvement of 22%. The observed *sensitivity* of the Haberman items in this sample (i.e., their successful prediction of students leaving the program) is 89%; they correctly predicted 31 out of 35 students who left. The observed *specificity* (i.e., their successful prediction of students remaining in and graduating from the program) is also 89%; they correctly predicted 64 out of 72 students who remained and graduated. Therefore, the authors recommend the use of the Haberman items as one measure of non-academic skills and dispositions to select teacher candidates. While the instrument is not a “silver bullet,” if used with caution and administered as trained, it can increase an EPPs likelihood of admitting candidates with the skills and dispositions for success in teacher education.

Although adding the covariates to Model 1 to create Model 2 provided statistical control and added slightly to predictive power, it should be kept in mind that they were nonsignificant. Nonsignificance, however, is more probably due to a lack of power than to an absence of effect, because, as noted in Results, the sample size was small relative to the number of predictors in Model 2, and the effects sizes for two of the covariates were very large. Minority status had an odds ratio of 5.45, meaning that minority students on average were more than five times as likely to leave the program as white students when Haberman scores are kept constant. Urban status has an odds ratio of 29.72, meaning that students from urban schools were almost 30 times as likely to leave as students from non-urban schools. The lack of significance precludes us from any conclusions about these predictors, but the lack of power and large effect sizes suggest that they are important factors to consider in future research on EPP retention. For example, when the “soft skills” measured by the Haberman are controlled, there may still be differences in the retention of urban and/or minority students, which itself implies that using the Haberman by itself will not help us reach the goal of diversifying the teaching workforce, and that other efforts at retention of these groups of students must be maintained, as well.

The estimated sensitivity and specificity of the Haberman as a tool for predicting program failure and success, respectively, in field-based urban teacher preparation programs similar to the one in this study, appears to be not only theoretically meaningful, but practically meaningful, as well. For example, if it cost the program \$60,000 in instructional and support costs over three years for each student, and 100 students were admitted, total costs would be \$6 million. If the program admitted 33 students who would eventually leave the program without graduating, \$1.98 million will have been spent in a less than optimal manner. However, if program administrators could reduce admissions of students who would not ultimately succeed by 67% (as in this sample), that would translate to a savings (or at least a return on investment) of about \$1.33 million, a sum that would seem to justify the additional costs (in terms of time, training, and money) of the Haberman applied with the model reported herein. However, readers are cautioned that the model may be overly precise given the limited sample on which it is based.

As standards for teacher education continue to increase, particular attention must be given to the ways in which we recruit, select, and prepare teachers for diverse communities in the United States. CAEP has made candidate quality, recruitment, and selectivity one of its main five standards. In meeting this standard, CAEP states that “the provider demonstrates that the quality of candidates is a continuing and purposeful part of its responsibility from recruitment, at admission, through the progression of courses and clinical experiences, and to decisions that completers are prepared to teach effectively and are recommended for certification” (CAEP, 2013). As indicated previously, “CAEP Commission recognizes the ongoing development of this knowledge base and recommends that CAEP revise criteria as evidence emerges” (CAEP, 2016a, p. 2). Studies such as this can inform these revisions to CAEP and related criteria. Employing proven selectivity measures, such as the Haberman Star Teacher Selection Interview, can assist teacher preparation programs in both demonstrating their purposeful admissions criteria, and in helping to increase retention of candidates

in the program. The authors caution, however, that the recommendation of employing selectivity measures that align with the context of their particular EPP, and the schools they serve, and that measure necessary non-academic skills of teachers is not a call for policy mandates for such instruments. Policy makers and accreditors must remain cognizant of the human side of teaching, and not mistake rigorous selection for silver bullets to be implemented by policy. EPPs and school districts must maintain autonomy in determining what skills and instruments are best aligned with their contexts and the needs of their communities. Selection can be rigorous without the creation of overly prescriptive policies.

Furthermore, as CAEP Standard 4 asserts, we must also ensure that program completers are fully prepared to teach effectively; this includes retention of beginning teachers. As indicated, preliminary program evaluations seem to indicate that graduates of the teacher preparation program involved in the current study are being retained at higher rates than the norm for beginning teachers. While 96% of teachers involved in this study were still teaching beyond their third year, other data from the program demonstrates that the five-year teacher retention rate for all graduates is 92%. Future studies are underway examining the relationship between the Haberman Star Teacher Selection scores, teacher retention, and student achievement in classrooms of program graduates. As we traverse deeper into the age of accountability, it is critical that teacher educators are able to provide credible evidence that the decisions we make are in the best interest of the children our graduates serve, and that we are continually striving for program improvement. Recruiting candidates with a commitment to serving all students, and utilizing selection instruments designed to identify effective teachers for diverse populations, comprise a critical first step in demonstrating our commitment and ability to recruit, select, and prepare effective teachers for our schools.

References

- American Association for Colleges of Teacher Education. (2013). The changing teacher preparation profession: A report from AACTE's Professional Education Data System. Retrieved from <http://aacte.org/news-room/announcements/aacte-releases-data-report-on-state-of-teacher-preparation.html>
- American Association for Colleges of Teacher Education. (2012). *Where we stand: Selectivity of educator preparation programs*. Retrieved from <http://aacte.org/research-policy/aacte-research-and-policy-statements/aacte-position-statements.html>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC; American Educational Research Association.
- Bastian, k. (n.d.). *Who teaches in North Carolina? Job placement rates for UNC system teacher education programs*. Education Policy Initiative at Carolina. Retrieved from <http://eqdashboard.northcarolina.edu/wp-content/uploads/2015/11/Job-Placement-Rates.pdf>
- Becker, B. J., Kennedy, M. M., & Hundesmarck, S. (2003, April). Communities of scholars, research, and debates about teacher quality. Paper presented at the meeting of the American Educational Research Association, Chicago.
- Berliner, D. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56(3), 205-213.
- Carpenter-LaGattuta, A. (2004). Challenges in multicultural teacher education. In F. Schultz (Ed.), *Annual editions: Multicultural education 04/05* (pp. 51-53). Guilford, CT: McGraw-Hill/Dushkin.

- Clark, S. K. (2010). Putting it to the test: Examining the use of standardized test scores to select teacher candidates into teacher preparation programs. *The World Federation of Associations of Teacher Education*. Retrieved from <http://www.wfate.org/conference10.html>
- CMU Career Services (2006). *Employment Survey Summary 2005-2006*. University of Central Michigan. Retrieved from https://www.cmich.edu/ess/academic_advising_assistance/career_services/AnnualReports/ar05_06.pdf
- Cochran-Smith, M. (2004). The multiple meanings of multicultural teacher education: A conceptual framework. In F. Schultz (Ed.), *Annual editions: Multicultural education 04/05* (pp. 26-35). Guilford, CT: McGraw-Hill/Dushkin.
- Conaway, B., Browning, L. & Purdum-Cassidy, B. (2007). Teacher candidates' changing perceptions of urban schools: Results of a 4-year study. *Action in Teacher Education*, 29, 20-31.
- Council for the Accreditation of Educator Preparation (2017). *Standard 4: Program impact components*. <http://www.caepnet.org/standards/standard-4>
- Council for the Accreditation of Educator Preparation (2016a). *Standard 3: Rationale*. Retrieved from <http://caepnet.org/standards/standard-3>
- Council for the Accreditation of Educator Preparation (2016b). *Standard 3: Components*. Retrieved from <http://caepnet.org/standards/standard-3>
- Council for the Accreditation of Educator Preparation. (2013). CAEP accreditation standards. Retrieved from https://caepnet.files.wordpress.com/2015/02/final_board_amended_20150213.pdf
- Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. National Commission on Teaching & America's Future. Kutztown Distribution Center Kutztown, PA.
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57(3), 300-314.
- Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, 61(1-2), 35-47.
- Darling-Hammond, L., & Baratz-Snowden, J. (2007). A good teacher in every classroom: Preparing the highly qualified teachers our children deserve. *Educational Horizons, Winter 2007*.
- Doll, D. (2009). *The correlation between the Haberman Star Teacher Selection Interview and the AASPA Interactive Computer Interview System – Urban*. (Doctoral dissertation). Retrieved from ProQuest. (1694502771).
- Ebmeier, H., Dillon, A., & Ng, J. (n.d.). Employment selection instruments—What we have learned from ten years of research. *Interactive Computer Interview System*. Retrieved from <http://www.people.ku.edu/~howard/ICIS.html>
- Eubanks, S., & Weaver, R. (1999). Excellence through diversity: Connecting the teacher quality and teacher diversity agendas. *Journal of Negro Education*, 68(3), 451-458.
- Furr, M. R., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Thousand Oaks, CA: Sage.
- Garmon, M. A. (2004). Changing preservice teachers' attitudes/beliefs about diversity? What are the critical factors? *Journal of Teacher Education*, 55(3), 201-213.
- Gay, G. (2000). *Culturally responsive teaching: Theory, research, and practice*. New York: Teachers College Press.
- Gay, G. (2004). Social studies teacher education for urban classrooms. In S. Adler (Ed.), *Critical issues in social studies teacher education* (pp. 75-95). Greenwich, CT: Information Age Publishing.

- Gay, G. (2010). Acting on beliefs in teacher education for cultural diversity. *Journal of Teacher Education*, 61(1-2), 143-152.
- Haberman, M. (2005a). *Star teachers: The ideology and best practice of effective teachers of diverse children and youth in poverty*. San Antonio, TX: The Haberman Educational Foundation.
- Haberman, M. (2005b). *Selecting and Preparing Urban Teachers*. Retrieved from <http://www.habermanfoundation.org/Articles?Default.aspx?=25>
- Haberman, M. (2009). *The rationale for preparing mature adults as teachers of diverse children in urban poverty*. Retrieved from <http://www.habermanfoundation.org/Articles?Default.aspx?=25>
- Haberman, M. (2011, December 15). Reliability of the Haberman Star Teacher Selection Interview. *Education Views*. Retrieved from <http://educationviews.org/reliability-of-the-haberman-star-teacher-selection-interview/>
- Haberman, M. & Post, L. (1998). Teachers for multicultural schools: the power of selection. *Theory into Practice*, 37(2), 96-104.
- Haberman Educational Foundation, Inc. (1994). *Star teacher selection interview training manual*. Houston: The Haberman Educational Foundation, Inc.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89.
- Hollins, E. (2012). *Learning to teach in urban schools: The transition from preparation to practice*. New York: Routledge.
- Ingersoll, R. (2003). The teacher shortage: Myth or reality? *Educational Horizons*, 81(3), 146-152.
- Ingersoll, R., Merrill, L., & Stuckey, D. (2014). Seven trends: The transformation of the teaching force, updated April 2014. *CPRE Report (#RR-80)*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children*. San Francisco: Jossey-Bass.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32, 465-491.
- Ladson-Billings, G. (2001). *Crossing over to Canaan: The journey of new teachers in diverse classrooms*. San Francisco: Jossey Bass.
- Ladson-Billings, G. (2006). It's not the culture of poverty, it's the poverty of culture: The problem with teacher education. *Anthropology and Education Quarterly*, 37(2), 104-209.
- Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children* (2nd ed). San Francisco, CA: Jossey-Bass.
- Metzger, S., & Wu, M. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, 78(4). Retrieved from <http://www.jstor.org/stable/40071150>
- Milner, R. (2010). What does teacher education have to do with teaching? Implications for diversity studies. *Journal of Teacher Education*, 61(1-2), 118-131.
- National Center for Education Statistics (2015). *Public School Teacher Attrition and Mobility in the First Five Years: Results From the First Through Fifth Waves of the 2007-08 Beginning Teacher Longitudinal Study First Look*. U.S. Department of Education. Retrieved from <https://nces.ed.gov/pubs2015/2015337.pdf>
- National Council for Accreditation of Teacher Education (2010). Transforming teacher education through clinical practice: A national strategy to prepare effective teachers. *Report of the Blue Ribbon Panel on Clinical Preparation and Partnerships for Improved Student Learning*. November, 2011. Retrieved from <http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3d&tabid=715>

- Pohan, C. (1996). Preservice teachers' beliefs about diversity: Uncovering factors leading to multicultural responsiveness. *Equity and Excellence in Education, 29*(3), 62-69.
- Ryan, P., & Alcock, M. (2002). Personal and interpersonal attributes in selecting teachers. *Action in Teacher Education, 24*(1), 58-67.
- Sandholtz, J. H., & Shea, L.M. (2012). Predicting performance: A comparison of university supervisors' predictions and teacher candidates' scores on a teaching performance assessment. *Journal of Teacher Education, 63*(1), 39-50.
- Sleeter, C. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of Whiteness. *Journal of Teacher Education, 52*(2), 94-106.
- Sleeter, C., & Thao, Y. (2007). Guest editors' introduction: Diversifying the teaching force. *Teacher Education Quarterly, 34*(4), 3-8.
- Stronge, J. H., & Hindman, J. L. (2006). *The Teacher Quality Index: A Protocol for Teacher Selection*. Alexandria, VA: Assoc. for Supervision and Curriculum Development.
- Teach for America (2016). *Is TFA right for me?* Retrieved from <https://www.teachforamerica.org/join-tfa/is-tfa-for-me>
- TNTP. (2016). *Selection criteria and indicators*. Retrieved from <http://tntp.org/teacher-talent-toolbox/explore/hiring>
- Tredway, L. (1999). The art of juggling: Preparing preservice teachers for urban schools. *Journal of Negro Education, 66*(3), 382-396.
- U.S. Department of Education (2000, November). Teacher quality initiative: Eliminating barriers to improving teaching. Retrieved from <http://www.ed.gov/teacherquality>
- University Career Center (2012). *2011-2012 Teacher Education Destiny Survey Report*. University of Kansas. Retrieved from <https://soe.ku.edu/sites/soe.ku.edu/files/docs/accreditation/Teacher%20Education%20Job%20Placement%2011-12.pdf>
- Villegas, A., & Lucas, T. (2002). Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education, 53*, 20-32.
- Waddell, J. (2011). Crossing borders without leaving town: the Impact of cultural immersion on the perceptions of teacher education candidates. *Issues in Teacher Education, 20*(2), 23-36.
- Waddell, J (2015). Recruiting, Supporting and Preparing Teachers for Urban Schools: The Institute for Urban Education. *Success in High Need Schools Journal, 11*(2).
- Waddell, J.H. (2013). Communities as critical partners in teacher education: The impact of community immersion on teacher candidates' understanding of self and teaching in urban schools. *Current Issues in Education, 16*(2), 15. Retrieved from <http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1198/505>
- Waddell, J. & Ukpokodu, O.N (2012). Breaking new grounds in recruiting and preparing diverse urban educators. *Multicultural Education, 20*(1), 15-22.
- Weiner, L. (1999). *Urban teaching: The essentials*. New York: Teachers College Press.
- Weiner, L. (2000). Research in the 90s: Implications for urban teacher preparation. *Review of Educational Research, 70*, 369-406.
- Zeichner, K. (2003). Pedagogy, knowledge and teacher preparation. In B. Williams (Ed.), *Closing the achievement gap: A vision for changing beliefs and practices* (2nd ed.) (pp. 99-114). Alexandria, VA: Association for Supervision and Curriculum Development.
- Zeichner, L. (2010). Rethinking the connections between campus courses and field experiences in college- and university-based teacher education. *Journal of Teacher Education, 61*(1-2), 89-99.
- Zeichner, K. (2011). *Rethinking college and university-based teacher education*. Unpublished manuscript, University of Washington, Seattle.

About the Authors

Jennifer H. Waddell

University of Missouri–Kansas City

waddellj@umkc.edu

Jennifer H. Waddell, Ph.D., is an Associate Professor of Teacher Education and Curriculum Studies in the School of Education at the University of Missouri–Kansas City where she also serves as the Director of the Institute of Urban Education and Sprint Foundation Endowed Professor of Urban Education. Dr. Waddell's research focuses on teacher preparation for diverse student populations, including policy and practice related to equity, diversity and urban schooling.

Jacob M. Marszalek

University of Missouri–Kansas City

<http://orcid.org/0000-0002-4319-2319>

marszalekj@umkc.edu

Jacob M. Marszalek, Ph.D., is an Associate Professor of Educational Research and Psychology in the School of Education at the University of Missouri–Kansas City, where he also serves as the Acting Director of the Urban Education Research Center, Dr. Marszalek teaches graduate courses in statistics, measurement, and research methods, and his research focuses on motivation and positive psychology, applications in test development and validity, and evaluation of educational outreach programs for women and minorities.

About the Guest Editors

Elena Aydarova

Auburn University

eza0029@auburn.edu

Website: <https://elenaaydarova.com>

Elena Aydarova is Assistant Professor of Social Foundations at the Department of Educational Foundations, Leadership, and Technology at Auburn University. Her interdisciplinary research examines the interactions between global social change and the work of teachers, teaching, and teacher education through the lens of equity and social justice. Her projects have explored teacher education reforms in Russia and the US, internationalization of education, teacher retention, as well as privatization of teacher preparation. She has recently completed a book manuscript "Teacher Education Reforms as Political Theater: Policy Dramas in Neoliberal Contexts." Throughout her career, Dr. Aydarova has taught in the United States, Ukraine, China, and the United Arab Emirates.

David C. Berliner

Arizona State University

berliner@asu.edu

David C. Berliner is Regents' Professor Emeritus of Education at Arizona State University. Dr. Berliner is a member of the National Academy of Education (NEA), the International Education Academy (IEA), a Fellow of the Center for Advanced Study in the Behavioral Sciences, and a past president of both the American Educational Research Association (AERA) and the Division of Educational Psychology of the American Psychological Association (APA). He is the recipient of awards for distinguished contributions from APA, AERA, and the

National Education Association (NEA). Dr. Berliner is co-author (with B. J. Biddle) of the best seller *The Manufactured Crisis*, co-author (with Ursula Casanova) of *Putting Research to Work*, co-author (with Gene Glass) of *50 Myths and Lies that Threaten America's Public Schools*, and co-author (with N. L. Gage) of six editions of the textbook *Educational Psychology*. He is co-editor of the first *Handbook of Educational Psychology* and the books *Talks to Teachers*, and *Perspectives on Instructional Time*. Professor Berliner has also authored more than 200 published articles, technical reports, and book chapters. He has taught at the University of Arizona, University of Massachusetts, Teachers College and Stanford University, as well as universities in Australia, Canada, The Netherlands, Spain, and Switzerland.

SPECIAL ISSUE
Navigating the Contested Terrain of
Teacher Education Policy and Practice

education policy analysis archives

Volume 26 Number 35

March 12, 2018

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank; SCOPUS, SOCOLAR (China).

Please send errata notes to Audrey Amrein-Beardsley at Audrey.beardsley@asu.edu

Join EPAA's Facebook community at <https://www.facebook.com/EPAAAPE> and Twitter feed @epaa_aape.

education policy analysis archives
editorial board

Lead Editor: **Audrey Amrein-Beardsley** (Arizona State University)

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Scott Marley, Iveta Silova, Maria Teresa Tatto** (Arizona State University)

Cristina Alfaro San Diego State University

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Jeff Bale OISE, University of Toronto, Canada

Aaron Bevanot SUNY Albany

David C. Berliner Arizona State University

Henry Braun Boston College

Casey Cobb University of Connecticut

Arnold Danzig San Jose State University

Linda Darling-Hammond Stanford University

Elizabeth H. DeBray University of Georgia

Chad d'Entremont Rennie Center for Education Research & Policy

John Diamond University of Wisconsin, Madison

Matthew Di Carlo Albert Shanker Institute

Sherman Dorn Arizona State University

Michael J. Dumas University of California, Berkeley

Kathy Escamilla University of Colorado, Boulder

Melissa Lynn Freeman Adams State College

Rachael Gabriel University of Connecticut

Amy Garrett Dikkers University of North Carolina, Wilmington

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Jacob P. K. Gross University of Louisville

Eric M. Haas WestEd

Julian Vasquez Heilig California State University, Sacramento

Kimberly Kappler Hewitt University of North Carolina Greensboro

Aimee Howley Ohio University

Steve Klees University of Maryland

Jaekyung Lee SUNY Buffalo

Jessica Nina Lester Indiana University

Amanda E. Lewis University of Illinois, Chicago

Chad R. Lochmiller Indiana University

Christopher Lubienski Indiana University

Sarah Lubienski Indiana University

William J. Mathis University of Colorado, Boulder

Michele S. Moses University of Colorado, Boulder

Julianne Moss Deakin University, Australia

Sharon Nichols University of Texas, San Antonio

Eric Parsons University of Missouri-Columbia

Amanda U. Potterton University of Kentucky

Susan L. Robertson Bristol University, UK

Gloria M. Rodriguez University of California, Davis

R. Anthony Rolle University of Houston

A. G. Rud Washington State University

Patricia Sánchez University of University of Texas, San Antonio

Janelle Scott University of California, Berkeley

Jack Schneider College of the Holy Cross

Noah Sobe Loyola University

Nelly P. Stromquist University of Maryland

Benjamin Superfine University of Illinois, Chicago

Adai Tefera Virginia Commonwealth University

Tina Trujillo University of California, Berkeley

Federico R. Waitoller University of Illinois, Chicago

Larisa Warhol University of Connecticut

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Terrence G. Wiley Center for Applied Linguistics

John Willinsky Stanford University

Jennifer R. Wolgemuth University of South Florida

Kyo Yamashiro Claremont Graduate University

archivos analíticos de políticas educativas
consejo editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editores Asociados: **Armando Alcántara Santuario** (Universidad Nacional Autónoma de México), **Jason Beech** (Universidad de San Andrés), **Ezequiel Gomez Caride** (Pontificia Universidad Católica Argentina), **Antonio Luzon** (Universidad de Granada), **Angelica Buendia** (Metropolitan Autonomous University), **José Luis Ramírez** (Universidad de Sonora)

Claudio Almonacid

Universidad Metropolitana de
Ciencias de la Educación, Chile

Miguel Ángel Arias Ortega

Universidad Autónoma de la Ciudad
de México

Xavier Besalú Costa

Universitat de Girona, España

Xavier Bonal Sarro Universidad
Autónoma de Barcelona, España

Antonio Bolívar Boitia Universidad
de Granada, España

José Joaquín Brunner Universidad
Diego Portales, Chile

Damián Canales Sánchez Instituto
Nacional para la Evaluación de la
Educación, México

Gabriela de la Cruz Flores
Universidad Nacional Autónoma de
México

Marco Antonio Delgado Fuentes
Universidad Iberoamericana, México

Inés Dussel, DIE-CINVESTAV,
México

Juan Carlos González Faraco

Universidad de Huelva, España

María Clemente Linuesa

Universidad de Salamanca, España

Jaume Martínez Bonafé

Universitat de València, España

Alejandro Márquez Jiménez

Instituto de Investigaciones sobre
la Universidad y la Educación,
UNAM, México

María Guadalupe Olivier Tellez,
Universidad Pedagógica Nacional,
México

Miguel Pereyra Universidad de
Granada, España

Mónica Pini Universidad Nacional
de San Martín, Argentina

Omar Orlando Pulido Chaves
Instituto para la Investigación
Educativa y el Desarrollo
Pedagógico (IDEP)

Paula Razquin Universidad de
San Andrés, Argentina

Miriam Rodríguez Vargas

Universidad Autónoma de
Tamaulipas, México

José Gregorio Rodríguez

Universidad Nacional de Colombia,
Colombia

Mario Rueda Beltrán Instituto de

Investigaciones sobre la
Universidad y la Educación,
UNAM, México

José Luis San Fabián Maroto

Universidad de Oviedo,
España

Jurjo Torres Santomé,

Universidad de la Coruña, España

Yengny Marisol Silva Laya

Universidad Iberoamericana,
México

Ernesto Treviño Ronzón

Universidad Veracruzana, México

Ernesto Treviño Villarreal

Universidad Diego Portales
Santiago, Chile

Antoni Verger Planells

Universidad Autónoma de
Barcelona, España

arquivos analíticos de políticas educativas conselho editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editoras Associadas: **Kaizo Iwakami Beltrao**, (Brazilian School of Public and Private Management - EBAPE/FGV, Brazil), **Geovana Mendonça Lunardi Mendes** (Universidade do Estado de Santa Catarina), **Gilberto José Miranda**, (Universidade Federal de Uberlândia, Brazil), **Marcia Pletsch, Sandra Regina Sales** (Universidade Federal Rural do Rio de Janeiro)

Almerindo Afonso

Universidade do Minho
Portugal

Alexandre Fernandez Vaz

Universidade Federal de Santa
Catarina, Brasil

José Augusto Pacheco

Universidade do Minho, Portugal

Rosanna Maria Barros Sá

Universidade do Algarve
Portugal

Regina Célia Linhares Hostins

Universidade do Vale do Itajaí,
Brasil

Jane Paiva

Universidade do Estado do Rio de
Janeiro, Brasil

Maria Helena Bonilla

Universidade Federal da Bahia
Brasil

Alfredo Macedo Gomes

Universidade Federal de Pernambuco
Brasil

Paulo Alberto Santos Vieira

Universidade do Estado de Mato
Grosso, Brasil

Rosa Maria Bueno Fischer

Universidade Federal do Rio Grande
do Sul, Brasil

Jefferson Mainardes

Universidade Estadual de Ponta
Grossa, Brasil

Fabiany de Cássia Tavares Silva

Universidade Federal do Mato
Grosso do Sul, Brasil

Alice Casimiro Lopes

Universidade do Estado do Rio de
Janeiro, Brasil

Jader Janer Moreira Lopes

Universidade Federal Fluminense e
Universidade Federal de Juiz de Fora,
Brasil

António Teodoro

Universidade Lusófona
Portugal

Suzana Feldens Schwertner

Centro Universitário Univates
Brasil

Debora Nunes

Universidade Federal do Rio Grande
do Norte, Brasil

Lílian do Valle

Universidade do Estado do Rio de
Janeiro, Brasil

Flávia Miller Naethe Motta

Universidade Federal Rural do Rio de
Janeiro, Brasil

Alda Junqueira Marin

Pontifícia Universidade Católica de
São Paulo, Brasil

Alfredo Veiga-Neto

Universidade Federal do Rio Grande
do Sul, Brasil

Dalila Andrade Oliveira

Universidade Federal de Minas
Gerais, Brasil