THE EVIDENCE BASED PRACTICE SELF-ASSESSMENT TOOL: EVIDENCE OF RELIABILITY AND VALIDITY

By Tiffany Sellars

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Educational Psychology

Northern Arizona University

December 2017

Approved:

John McClure, Ph.D., Co-Chair

Trina Spencer, Ph.D., Co-Chair

Lena Gaddis, Ph.D.

Katherine Mahosky, Ed.D.

Abstract

THE EVIDENCE BASED PRACTICE SELF-ASSESSMENT TOOL:

EVIDENCE OF RELIABILITY AND VALIDITY

This study examined the factor structure, internal consistency reliability, content validity index, known-groups validity, and feasibility, acceptability, and usability of the Evidence Based Practice Self-Assessment Tool (EBPSAT). Survey data were collected from a sample of 291 teachers, administrators, and school-based related service providers. The participants were largely white (82%) and female (77%) who completed the EBPSAT using Survey Monkey. Exploratory factor analysis (EFA) was performed to examine the underlying factor structure of the instrument. The EFA resulted in a two-factor structure. The content validity index (CVI) was calculated, resulting in adequate overall evidence of content validity. One-way ANOVA was conducted to determine the EBPSAT's ability to differentiate between groups. Implications of these results, as well as suggested changes to the EBPSAT, including item elimination and/or revisions, are provided.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

This work would not have been possible without the guidance and support of my committee members, the encouragement of my family, and the moral support and feedback from my fellow interns at the Louisiana School Psychology Internship Consortium. Specifically, I must express my deepest gratitude to Dr. Trina Spencer, who first introduced to me to evidence-based practice when I was a graduate research assistant at the Institute for Human Development. Though our relationship got off to a rocky start (when I failed to show up for my first week of work), Dr. Spencer has had lasting impact on my life, both personally and professionally. It was her passion for research and the use of evidence in improving the lives of children that has significantly influenced my own research interests. Dr. McClure served as an excellent and patient chair and an invaluable resource as he has worked with me in completing this dissertation. Dr. Gaddis has also been an important professional role model and an integral part of my NAU experience. Finally, Dr. Katherine Mahosky has been essential in providing me with a fresh perspective and insightful feedback.

In addition to my committee members, I'd also like to thank my colleague Stephanie Heron, who tolerated many long distance phone calls, acting as sounding board and good friend throughout this process; my fellow intern, now Dr. Erin McClure, who reminded me to take time for myself; and Dr. Whitney Kleinert who reminded me not to take too much time for myself. Finally, genuine thanks to my family who provided both unwaivering support and frequent reminders to get back to work.

Chapter 1

Introduction

The *evidence-based practice* (EBP) movement in education began with the *No Child Left Behind Act* (NCLB) of 2001, which called for the use of educational interventions and programs established through "scientifically-based research" (US Department of Education, 2003, p. iii), referring to the "rigorous, systematic, and objective methods of science to examine and validate instructional procedures" (Yell, Shriner, & Katsiyannis, 2006, p.11). The following year, the Institute of Education Sciences (IES) was created as part of the *Education Science Reform Act* of 2002, to "promote objective, high quality research in education" (Hanley, Chambers, & Haslam, 2016, p. 288). The importance of EBP in education was further emphasized by the 2004 reauthorization of the *Individuals with Disabilities Education Act* (IDEA), which was aligned with the provisions outlined in NCLB (Yell et al., 2006). Finally, the 2015 reauthorization of NCLB, renamed the *Every Student Succeeds Act* (ESSA), explicitly defines the term "evidence-based" and outlines specific criteria for research to be considered as such.

Definitions of Evidence-based Practice

The term *evidence-based practice* (EBP) has been defined across a variety of disciplines including: medicine, mental health, social work, speech language pathology, occupational therapy, physical therapy, and education. Having originally appeared in the medical literature in the 1990's, EBP has since been adapted to reflect philosophical differences in terms of how practices are identified and utilized in each discipline (Juniel, 2015). While the primary objective in defining and applying EBP is to improve outcomes for the populations being served, its conceptualization varies by discipline.

In the term evidence-based practice, the use of the word "practice" can apply to both the

professional activities of an individual, much as an attorney *practices* law, and to the "specific methods or techniques used by a professional" (Spencer, Detrich, & Slocum, 2012, p. 129). Thus, in the broad definition EBP is a verb, referring to the process of making decisions regarding appropriate techniques, while the much narrower interpretation defines it as a noun, referencing specific techniques. For the purpose of the current study, *evidence-based practice* will refer to the broader definition of EBP as a decision making process, and the term *empirically supported treatment* or *intervention* will be used to mean specific methods or techniques used by professionals.

Medicine. EBP initially appeared in the literature in the early 1990's as *evidence based medicine* (EBM), which was defined as, "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett, Rosenberg, Haynes, & Richardson 1996). Gordon Guyatt M.D. (1991) illustrated this new approach through a case study, in which a medical intern suspects a patient of having iron deficiency anemia. Following the previous approach to determining the diagnosis, the intern would have ordered standard tests as direct by the attending physician and proceed based on the results of the tests.

Following the EBM approach, the intern would examine diagnostic properties of the tests before conducting a brief literature search, using that information to inform her next steps. This approach integrates current evidence into the decision-making process. According to Guyatt (1991), evidence-based medicine "uses additional strategies, including quickly tracking down publications of studies that are directly relevant to the clinical problem, critically appraising these studies, and applying the results of the best studies to the clinical problem at hand" (p. A16).

Psychology. In 2005, the American Psychological Association appointed the APA

Presidential Task Force on Evidence-Based Practice (Task Force). The Task Force was charged with completing three objectives: (1) consider how a range of research evidence should be integrated into the practice of psychology, (2) to "articulate and explicate" (p.4) the role of clinical expertise in decision making, and (3) to "articulate and explicate" (p.4) the role of patient values in decision making. The Task Force agreed upon the following definition of evidence-based practice, "Evidence-based practice in psychology is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (p. 5). The Task Force report stated that the purpose of evidence-based practice in psychology "is to promote effective psychological practices and promote public health by applying empirically supported principles of psychological assessment, case formulation, therapeutic relationship, and intervention" (p. 5).

Social Work. In the field of social work, EBP is defined as, "the explicit and judicious use of current best practice in making decisions regarding the welfare of those in need" (Sheldon & Chivers, 2002, p. 26). More explicitly, EBP is "a process that allows social workers to identify, evaluate, and apply evidence relevant to a client's issues to practice decisions" (Jenson, 2007 as cited in Diaz & Drewery, 2016). However, the literature indicates that the definition is not effectively taught or well understood (Drisko & Grady, 2015; Gray & Schubert, 2012; Wike et al., 2014). Further, Wike and colleagues (2014) contend that the definition of EBP assumes that practitioners have the skills necessary to evaluate and integrate research evidence, which may not be the case.

Education. Prior to the introduction of the *No Child Left Behind Act* (2001), the field of education lacked established processes for identifying and implementing empirically supported interventions. In fact, Slavin (2002) contends that, at the beginning of the 21st century, education

was just entering the 20th century regarding the integration of scientific research into practice. He further states that, "The use of randomized experiments that transformed medicine, agriculture, and technology in the 20th century is now beginning to affect educational policy" (p. 15). *NCLB* represents the first time that the use of scientific research in making educational decisions was featured in national education legislature (Spencer et al., 2012).

In education, the term *evidence-based practice* has primarily been used to refer to specific interventions or programs that have strong research support (Spencer et al, 2012). For example, the *Every Student Succeeds Act* (ESSA), the 2015 reauthorization of *NCLB*, explicitly defines "evidence-based" as,

> …an activity, strategy, or intervention that: (i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on— (I) strong evidence from at least one well-designed and well-implemented experimental study, (II) moderate evidence from at least one well-designed and well-implemented quasi-experimental study, or (III) promising evidence from at least one well-designed, and well-implemented correlational study with statistical controls for selection bias; or (ii) (I) demonstrates rationale based on high quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes and (II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention (Fleischman, Scott, & Sargard, 2016, p. 5).

Applying EBP in Education

As the term *evidence-based practice* originated in the medical field, there are some reasonable challenges in applying it to the field of education. For example, teacher education programs offer little to no foundation in conducting or evaluating research; whereas, an

understanding and appreciation of science is inherent in medicine (AlGhamdi et.al, 2013; Jahan, Maqbali, Siddiqui, & Zadjali, 2005). Further compounding the problem, research indicates that teachers, as a group, are more resistant to change than individuals in other fields (Vanlommel, Vanhoof, & Petegem, 2016). Finally, randomized controlled trials, which have long been considered the "gold standard" in providing research evidence, are difficult to conduct in education. There is also an "added complexity in education that classrooms consist of many individuals, who may have different individual experiences and attitudes but also develop a collective response to what teachers do" (Wrigley, 2016, p. 241), meaning that individual classes or students may have different responses to instructional practices, and responses may vary according to environmental differences. Essentially, strategies that work on Monday afternoon may not be effective on Friday morning.

ESSA's tiered approach to evaluating research evidence allows for more flexibility in determining which approaches are supported by evidence and the strength of the evidence. The added flexibility will permit decision-makers to consider the acceptability of evidence within the context of individual schools (Fleischman et al., 2016). However, greater flexibility places greater emphasis on *clinical judgment* for selecting appropriate empirically supported interventions.

The APA Presidential Task- Force's (2005) definition of *clinical expertise* included eight competencies to promote positive outcomes. The competencies are as follows: (1) assessment and diagnostic judgment; (2) clinical decision making, implementation, and progress monitoring; (3) interpersonal expertise; (4) self-reflection and skill acquisition; (5) evaluation and use of research evidence in practice; (6) understanding the influence of cultural and individual factors; (7) seeking available resources; and (8) having a logical rationale for strategy selection. Though

the definition is intended for psychologists, it can be translated for use in education with few alterations. Educators would ideally be competent in each of these areas or, at the very least; they should be familiar with the resources that are available to them.

Intervention Resources for Educators

What Works Clearinghouse. To help facilitate the process of selecting empirically supported interventions, the Institute of Education Sciences created the *What Works Clearinghouse* (WWC), a database of more than 700 publications and catalogs and over 11,000 reviewed research studies that are rated by their degree of evidence. Jerome D'Agostino, a professor of educational studies at Ohio State University, recently led a meta-analysis of WWC's reviews, which identified only 29 interventions with significant effects out of over 10,000 (Sparks, 2016). D'Agostino said, "If you look from 10,000 feet at education interventions, you can almost count on your hand the number of interventions that have truly scaled and established (themselves)" (Sparks, 2016, p. 8). The results of this meta-analysis further emphasize the necessity of a flexible system of identifying intervention "evidence".

The Institute of Education Sciences recently updated the WWC website to allow individuals to search for studies based on school characteristics and demographics (Sparks, 2016). The new format should help school representatives find interventions that have demonstrated significant effects in specific contexts; thus making the decision-making process of EBP user friendly.

Results First Clearinghouse Database. The Results First Clearinghouse database was created in 2011 by the Pew-MacArthur Results First initiative. The database was developed to assist policy makers in identifying empirically supported programs for making "data-driven budget decisions" (Haskins, 2015, p.1). It integrates eight databases, including What Works

Clearinghouse, that evaluate the research evidence for studies in a variety of disciplines (e.g., mental health, social policy, health care, and criminal justice). The database uses a color-coded system for categorizing research evidence from green (highest rated) to red (negative effects). However, within the past twelve months, two of the eight databases have shut down, citing a lack of funds.

Purpose of the Study

The purpose of the current study was to investigate the psychometric properties of the newly developed Evidence Based Practice Self-Assessment Tool (EBPSAT) (Spencer, 2016). Currently, a measure assessing attitudes toward and use of evidence-based practice specifically in education does not exist. The participants were teachers, support staff, and administrators who were contacted through education listservs. Potential participants received an email with instructions and a link to the electronic survey. To increase sample size, snowballing methods were utilized via social media.

The EBP Self-Assessment Tool

The Evidence-Based Practice Self-Assessment Tool is a survey instrument designed to address the extent to which educators and educational organizations implement evidence-based practice as a decision-making process. It includes 20 items across three domains: *Organization*, *Individual*, and *Practice*. The *Organization* domain includes items assessing the respondent's perceptions of evidence-based practice implementation and support at the organization in which the respondent is employed, while the *Individual* domain assesses personal perceptions and use of evidence based practice. The items in the *Organization* and *Individual* domains require Likert-type responses on a scale from 1 (strongly agree) to 4 (strongly disagree), or 0 (don't

know). The *Practice* domain includes five items to examine the frequency of an individual's engagement in EBP activities such as reading and discussing current research.

Research Questions

R_1: What is the underlying factor structure of the Evidence Based Practice Self-Assessment Tool?

R_2: What is the internal consistency reliability of the Evidence Based Practice Self-Assessment Tool?

R_3: What is the content-validity evidence of the Evidence Based Practice Self-Assessment Tool?

R_4: What is the construct-validity evidence of the Evidence Based Practice Self-Assessment Tool?

R_5: What is the evidence of usability, acceptability, and feasibility of the EBPSAT according to educators and school-based related service providers?

Delimitations

The delimiting factors in the current study include the method of recruitment, the voluntary nature of participation, and the lack of criterion measures for comparison. The study will include teachers, administrators, and related service providers working in schools. Since participants will be recruited via a department of education listserv, characteristics of the sample may be affected. This method of recruitment could also potentially yield unequal representations of participants from various subgroups within schools (i.e. teachers vs. paraprofessionals vs. administrators). The voluntary nature of participations also serves as a delimiting factor as characteristics of individuals who voluntarily complete a survey instrument may differ from the general population. Finally, the lack of reliable and valid criterion measures of EBP use serves as a delimiting factor, which could impact the ability of the current study to determine the

construct validity of the EBPSAT.

Significance of the Study

Education research is described as scientific studies *on* education and *for* education (Vanderlinde & van Braak, 2010) that has the potential to guide practice and foster school improvement. Logically, it would then follow that such research would be used to inform decisions regarding educational practices. The research-to-practice gap is a phrase used to denote the discrepancy between educational research and the practices and interventions being employed in schools (Vanderlinde & van Braak, 2010). The nature of the gap has been a subject of dispute among researchers, practitioners, and policy. Some question the assumption that there is sufficient evidence of intervention effectiveness to utilize such empirically supported strategies in schools, especially for low-income communities (Frazier, Formoso, Birman, & Atkins, 2008); while others argue that the gap is a result of differing objectives between researchers and educators. Researchers aim to seek new knowledge, while educators seek solutions to operational problems in the classroom (Frazier, Formoso Birman, & Atkins, 2008).

To attempt to bridge the research-to-practice gap and to improve student outcomes, legislation such as *NCLB, IDEA 2004,* and *ESSA* mandate the use of "scientifically-based research" and "evidence-based practice" in educational decision-making (US Department of Education, 2003). Although new legislation may seem a rational solution to the problem, there is a disparity between making a law and realistically implementing its directives. The EBPSAT is designed to assist educational organizations in identifying areas of weakness within their organization regarding EBP implementation.

Identifying and eliminating barriers that prevent decision-makers from being able to use research effectively to inform practice is a step toward greater use of empirically supported

interventions in schools and, ultimately, greater student outcomes. This study was an exploratory study to investigate the psychometric properties of a new instrument that has the potential to examine school personnel's perceptions toward and use of EBP in order to inform goals and guide professional development activities.

Definition of Terms

Empirically Supported Interventions- Specific programs or interventions used by educators that have strong research evidence to support them.

ESSA- The Every Student Succeeds Act of 2015, federal reauthorization of the Elementary and Secondary Education Act originally passed in 1965, previously titled *No Child Left Behind Act*, 2001. This law introduced a tiered system of evaluating research evidence for interventions.

EBP- Evidence Based Practice, a decision making process that is informed by three sources: 1) best available research evidence, 2) clinical expertise, and 3) client and family characteristics, values, and preferences. Importantly, the word "practice" refers to all professional activities of an individual and not a single intervention or program.

IDEA 2004- The Individuals with Disabilities Education Improvement Act, federal legislation passed in 2004, which featured a provision to allow for the use of
Response to Intervention in the diagnosis of specific learning disabilities
(Individuals with Disabilities Improvement Act, 2004).

Interventions- Instructional techniques employed by educators to achieve a specific set of outcomes.

NCLB- The No Child Left Behind Act of 2001, federal reauthorization of the
Elementary and Secondary Education Act originally passed in 1965, contributed to the movement towards Response to Intervention through its focus on scientifically based practice and accountability for student academic achievement with additional emphasis on these principles as applied to reading through the Reading First section of NCLB (Mellard & Johnson, 2008; No Child Left Behind Act, 2002).

Scientifically-based Research- Defined by NCLB as, research that involves the application of

rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs

Chapter 2

Review of the Literature

The evidence-based practice movement in education began with landmark legislature such as the No Child Left Behind (NCLB) Act of 2001, which was soon followed by the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA), and most recently, the Every Student Succeeds Act (ESSA) of 2015. While these laws emphasized the use of research in making educational decisions, they did not provide a blueprint for doing so, making it difficult for educators to comply. Difficulties with EBP implementation can also be attributed to the well-known *research-to-practice gap,* which refers to the "significant and persistent gap" (Cook & Cook, 2013, p. 71) between the research evidence and the professionals working in the field.

Within the literature related to the research-to-practice gap, there are competing theories regarding its cause. Some argue that the gap is caused by teachers' general distrust of educational research (Cook, Cook, & Landrum, 2013), while others claim that the blame lies with the accessibility of research findings to practitioners, or in the research evidence itself. While, others still, cite the complex academic language as the reason for teachers' aversion to research evidence. The research evidence regarding teachers' perspectives and use of EBP primarily relies on qualitative methods as there are few instruments to measure the use of EBP in schools that encompass the complexity of the construct. The availability of such an instrument has the potential to inform professional development for teachers, administrators, and related service providers in the school setting; thus, ultimately positively impacting student outcomes.

A Closer Look at NCLB, IDEIA, and ESSA

NCLB (2001) was signed into law as, the most recent reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965. As the first federal law to address facets of

education that had previously been left to state and local decision makers, ESEA was considered groundbreaking (Duffy, Giordano, Farrell, Paneque, & Crump, 2008). It introduced Title I, which directed federal funds to local education agencies (LEAs) to "improve educational opportunities for disadvantaged children"(Duffy et al., 2008, p.55), while giving the federal government a role in public school policy decisions. However, the ambiguity of ESEA led Congress to amend the law four times between 1965 and 1980.

No Child Left Behind. While the intent of NCLB, which was to, "To close the achievement gap with accountability, flexibility, and choice, so that no child is left behind" (2001) was aligned with the 1994 reauthorization of ESEA, the Improving America's Schools Act, NCLB included provisions for assessments tied to accountability that were lacking in the previous authorization (Yell et al., 2006b). In order to keep their federal funds, schools now had to provide documentation of student success. Schools also had to employ "high quality" teachers, according to a federal definition of "high quality", which included requirements such as holding a bachelors degree and passing state licensure exams ("Overview", 2008). Further, these "highly qualified" teachers were now required to use "scientifically based research" (No Child Left Behind, 2001) in their practice.

Within the text of NCLB, the term "scientifically based research" is referenced over 100 times (Whitehurst, 2002). The law defines scientifically based research as, "research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs" (No Child Left Behind, 2001). In order to meet NCLB criteria for scientifically based research, programs must, "(1) use systematic, empirical methods based on experimental design; (2) involve rigorous data analyses to test hypotheses; (3) rely on measurements that produce valid data; and (4) be accepted by a

peer reviewed journal or panel of experts" (Trybus, 2007, p. 5). The impact of NCLB has been felt in every public school across the United States, with implications for both general and special education students.

Individuals with Disabilities Education Improvement Act (IDEA). The Education of All Handicapped Children Act (PL 94-142) became a federal law in 1975. It required the provision of free and appropriate education (FAPE) suited to students' individual needs. Prior to its passage, an estimated one million children were excluded from educational opportunities and another three million had very limited exposure (Detrich, Keyworth, & States, 2016). PL 94-142 had four primary goals: (1) to improve identification and education of students with disabilities; (2) to protect the rights of students and their families; (3) to assist states in their efforts to provide an education to students with disabilities; and (4) to assess the effectiveness of its efforts (US Department of Education, 2010). To further attempt to achieve goals for providing education to individuals with disabilities, several amendments were added to PL 94-142 between 1975 and 2000, including: that state agencies provide services to children from birth, rather than three years of age and a provision of transition services for high school students (US Department of Education, 2010). The 1990 amendments changed the name of the law to the *Individuals with Disabilities Education Act.*

The 2004 reauthorization of IDEA, renamed the Individuals with Disabilities Education Improvement Act (IDEIA), made important changes to the law, which closely aligned it with NCLB (Yell, Katsiyannis, & Shriner, 2006a). IDEIA was aligned with NCLB in three major ways including requiring that (1) special education teachers be highly qualified; (2) students with disabilities participate in statewide assessment; and (3) special education practices be research-based (Yell et al., 2006a). The 2004 amendments also allowed states to "employ a response-to-

intervention framework to consider students' response to scientific, research-based intervention when identifying students with specific learning disabilities" (US Department of Education, 2010, p. 10), further emphasizing the need for the use of evidence-based practice in schools.

Every Student Succeeds Act (ESSA). Signed into law in December of 2015, ESSA reauthorized the Elementary and Secondary Education Act and replaced NCLB. There are two key elements of the new legislation that stand out: a shift away from federal mandates toward increased state and local authority and an emphasis on evidence-based practices for school improvement. In regards to EBP, ESSA established three tiers of acceptable evidence: Tier I representing "strong evidence"; Tier II representing "moderate evidence; and Tier III indicating "promising evidence" (Fleischman et al., 2016).

NLCB's definition of "scientifically-based research" primarily referred to randomized control trials (RCTs) and certain studies employing quasi-experimental designs (QEDs), whereas, ESSA's tiered system allows for more flexibility in determining the "evidence" in evidence-based practice. Each of the tiers outlines the type of research study that qualifies. For example, Tier I includes "at least one well-designed and well-implemented experimental study" (Fleischman et al., 2016, p. 5). Tier II includes QEDs and Tier III refers to correlational research. This system may help school decision-makers select programs and strategies that suit school needs and local contexts. Whereas NCLB did little to promote a unified approach to evidence-use in schools, ESSA has the potential to improve educational practices through unified language and a wider range of "acceptable" approaches (Fleischman et al., 2016).

The Achievement Gap in Education

The overarching purpose of implementing research evidence in education is to improve student achievement through the use of *effective* interventions. However, the majority of

students' scores on the National Assessment of Educational Progress (NAEP) remain below 'proficient' in reading and math. Specifically, in 2015, 36% of fourth graders scored at or above proficient in reading, while 40% scored at or above proficient in math (US Department of Education, 2015). The data are even more discouraging for Black (18% reading; 13% math) and Hispanic (21% reading; 26% math) students. Like other countries, the United States, has student achievement profiles that are ethnically and economically stratified (Timperley & Parr, 2007).

It was NCLB's requirement that student achievement data be disaggregated by subgroup, which highlighted the "glaring and profound inequities" that are found in America's education system (Noguera, 2009, p. 61). Since the magnitude of the problem was revealed, a number of attempts have been made to reform schools and reeducate teachers, since research indicates that teachers have the greatest impact on student achievement. However, many of these efforts generally result in "relatively small and typically unreliable achievement gains" (Timperley & Parr, 2007, p. 90).

Critics of NCLB claim that its forced accountability measures mistake measuring schools for fixing them. While others further argue that it fails to address the complexities underlying why ethnic minorities and students from low socioeconomic backgrounds often struggle in school (Sherman, 2008). The point is that there are a variety of factors contributing to the achievement gap including: low teacher expectations, racist attitudes among staff, and lack of choice for parents in selecting schools. Additionally, NCLB assumes that all schools have the capacity for positive change, which may not be accurate (Sherman, 2008). The realities of the achievement gap emphasize the necessity of EBP in schools. The EBP framework encourages educators to select empirically validated interventions and teaching strategies based upon their current context as well as student characteristics. However, the research suggests that there are

barriers preventing the use of research evidence in schools.

The Research-to-Practice Gap

According to Cook et al. (2013), "The argument is simple: by researchers clearly identifying practices shown by trustworthy bodies of research to be effective, practitioners can know and implement what really works, thereby improving student outcomes" (p. 164). Unfortunately, as simple as it may seem to implement research in practice, there is a multitude of factors contributing to the "significant and persistent gap" (Cook & Cook, 2013, p. 71) between the research evidence and the professionals working in the field, one of which is that many teachers do not hold education research in very high esteem. Similarly, researchers critique teachers' reliance on experiential knowledge and failure to use the most current evidence to inform their decisions (Gore & Gitlin, 2004). There is also the problem of accessibility—meaning that teachers are often unable to locate and use the research, prevented from doing so by lacking subscriptions to peer-reviewed journals or by the academic language used to write the articles. Additionally, teachers may not have the time or the training to be consumers of research.

Teachers' perceptions of research. Gore and Gitlin (2004) examined teachers' perspectives of research, using interview and questionnaire data. The participants included 85 pre-service teachers and 147 practicing teachers in the United States and Australia. The researchers found that teachers were primarily concerned with the practical utility (or lack thereof) of education research. Many participants reported that did not find much value in research because it does not directly apply to their everyday experiences. One such participant stated that, "Theory is not relevant to someone working in the classroom" (p. 40). Others went so far as to provide scathing remarks, referring to academic research as "appalling", "laughable", and "bunk" (Cook et al., 2013, p.39).

Additionally, teachers not only want research that is applicable in a classroom; they want it to apply to their specific contexts. Since many teachers believe their particular context to be unique, they question the usefulness of research in addressing challenges they face in the classroom (Cain, 2016). Gore and Gitlin (2004) state that, "[Teachers] doubt the relevance of research produced outside their local contexts and convey a sense that research does not and cannot, even though they want it to, answer questions that may be specific to a single classroom" (p. 41). This perspective can partially be explained by the lack of movement of teachers between classrooms and schools. This isolation prevents teachers from observing the contextual similarities between classrooms and limits the types of knowledge they might consider when making decisions.

Landrum et al. (2002) examined perceptions of the trustworthiness, usability, and accessibility of information from different sources according to 127 experienced teachers (67 general education and 60 special education). The sources of information included: (1) experienced teachers; (2) conferences, workshops, or in-service presentations; (3) college or university coursework; and (4) professional journals. They found that participants consistently rated information from professional journals as less trustworthy, usable, and accessible than information from colleagues and conference presentations. There were no significant differences associated with years of teaching or between general and special education teachers. It may be that teachers value information and insight from their colleagues over research findings because they perceive that their fellow teachers can understand the context and their particular challenges.

To examine teachers' tendency to trust information provided by their colleagues, Landrum et al. (2002) presented two groups of teachers with three procedures for teaching

reading, written in one of two formats: a data-based perspective or a personal anecdotal perspective. The data-based perspective was written like a research article and referenced fictitious studies, while the personal perspective was written as a personal account from an experienced teacher. After reading each procedure, participants rated it on factors related to usability (i.e. I would use the procedure described, the teaching procedures are easy to use, etc.). They found that teachers had "relatively positive attitudes" (Landrum et al., 2002, p. 34) toward both presentation formats, which the authors attributed to the fact that teachers are predisposed to consider teaching techniques about which they have read.

In a more recent study in the UK, Procter (2015) found that teachers in their sample placed a high value on research practices. However, there were significant gaps between what teachers value and what occurs in practice, both at the individual and the school level. Using a dual Likert scale questionnaire, researchers investigated the gap between values and practice by asking participants to rate how often they engage in different practices alongside how highly they value that particular practice. For example, engaging in research-focused discussion received a value score of 74.6%, but a practice score of only 39.7%, resulting in a gap of 34.9%. These findings suggests the presence of environmental constraints, which prevent teachers from actively engaging with research regardless of their receptivity to it. Though it should be noted that, of the 156 survey respondents, 28 (19.8%) were studying for a Master's Degree, indicating that the sample may not be representative of the general population regarding research-related values.

However, perceptions of research seem to vary across educators in general. Among a group of ten new special education teachers, Jones (2009) identified four as "definitive supporters" of research, three as "cautious consumers", and three as "critics". Similarly,

Vanderlinde and van Braak (2010) conducted focus group interviews of teachers, school leaders, researchers, and intermediaries (defined as individuals responsible for translating and/or distributing research findings), which found teachers to be more skeptical of research than are school leaders. The study results also indicated that teachers perceive the gap between research and practice to be larger than do other groups. Whereas, school leaders were more likely to report reading research articles and attempting to integrate evidence into practice, teachers were skeptical of the usefulness of research evidence in practice. These studies suggest that the variability in educators' views of research is influenced by a range of factors including values, experience, role, and seniority (Cain, 2016; Gore & Gitlin, 2004; Huat See, Gorard, & Siddiqui, 2016).

Challenges of accessibility. It is often mistakenly assumed that the relationship between research and practice is linear. Researchers provide the evidence, which is then translated into practice by the practitioner (i.e. teachers) through dissemination, defined as the "translation and distribution of research findings to practitioners" (Vanderlinde & van Braak, 2010, p. 302). However, this relationship is infinitely more complex.

In investigating teachers' perceptions of educational research, a common theme has emerged in the literature; many teachers claim that, though they may have the desire to engage with research-evidence, they simply do not have access to it (Levin, 2004; Neal, Neal, Kornbluh, Mills, & Lawlor, 2015; Rey & Gaussel, 2016; Vanderlinde & van Braak, 2010). Vanderlinde and van Braak (2010) hypothesize that this is due, in part, to the conflicting objectives of researchers and practitioners. The authors state that, "the practitioner asks for new solutions to operational problems while the researcher seeks new knowledge," (p. 301) which contributes to a lack of reciprocal communication between researchers and practitioners (Neal et al., 2015).

Further complicating matters, the "publish or perish" phenomenon in academia provides little incentive for researchers to submit articles to "practitioner journals" rather than peer-reviewed academic journals (Neal et al, 2015; Vanderlinde & van Braak, 2010). In fact, doing so could impact future tenure decisions since practitioner journals are perceived as "less scholarly" (Vanderlinde & van Braak, 2010, p. 302). Thus, relevant research findings remain out of teachers' grasp, as most teachers do not have subscriptions to peer-reviewed academic journals.

Neal et al. (2015) advocate for the use of research "brokers", meaning "persons or organizations that facilitate the creation, sharing, and use of knowledge" (Meyer, 2010, p. 119, as cited in Neal et al. 2015). Brokers can serve the function of bringing the research to the practitioners in a format that they can easily understand. According to Cain (2016), teachers, when given the opportunity to engage with research-evidence, have difficulty comprehending the complex academic language used to write many scholarly articles. Huat See et al. (2016) found evidence of this when teachers demonstrated the ability to apply research findings, but only at a surface level. The authors state that, "…academic papers are not written for practical application and not specifically meant for practitioners. Such papers do not usually give detailed descriptions of interventions and how they are to be implemented" (p. 69).

Not only do practitioners struggle with the academic language of research articles, they also have difficulty comprehending the results and implications of the studies. Teachers are not trained to critically evaluate research findings and may not be able to differentiate between effective and ineffective treatments or high or low quality methods (Emmons, Keefe, Moore, Sánchez, & Neely, 2009). In a study of the use of research-based practices in special education, Burns and Ysseldyke (2009) found that special educators frequently engage in practices that research has shown to have low effect sizes such as social skills training (mean effect size of .21)

and modality training (mean effect size of .14). The researchers also found that teachers rated two of the most effective practices (according to research), mnemonic strategies and applied behavior analysis, less favorably than other, less effective, strategies.

It is evident that limited research accessibility contributes to the research-to-practice gap. Practitioners often do not have access to the journals in which the evidence is published. When they are able to physically access the research, they are prevented from comprehending the key points because of the unfamiliarity of the methods and academic language used. Finally, as practitioners are not generally trained to critically evaluate the research, they may misinterpret the results. Research brokers could potentially help with this problem, however, Neal et al. (2015) found there to be few brokers in the school setting.

The 'evidence' in EBP. In addition to perceptions of research and issues with accessibility, there are challenges inherent in the research evidence itself. What constitutes "evidence" has long been a point of contention among education researchers, policy makers, and practitioners. Though randomized control trials (RCTs) are ideal in the medical field, many argue that they do not translate well to education for a variety of reasons and too little attention is given to other research methodologies such as single-case design (SCD), quasi-experimental, and correlational designs. Danforth (2006) argues that the US Department of Education has, " taken a 'hard science' stance on what counts as knowledge, calling for experimental designs that are more common to medical research than to educational inquiry" (p.338).

Randomized control trials. RCTs have long been considered the "gold standard" of research evidence. With the original aim of reducing bias and enhancing the accuracy of clinical experimentation, they have reshaped knowledge production in the medical field and are considered by many to meet the highest standards of rigor (Bothwell, Greene, Podolsky, &

Jones, 2016; Slavin, 2002). Though, when RCTs were first introduced they received mixed reviews. Critics were concerned with the ethics of withholding treatment from control groups, while proponents argued that the methodology could determine if new treatments and interventions were better than the standard of care control groups received (Bothwell et al., 2016).

In education, some argue that RCTs are not always the most appropriate means of generating research evidence. According to Norman (2004), "the randomized control trial, whatever its virtues in evaluating drugs, is simply the wrong way to go in looking at curriculum interventions" (p. 257).Further arguments against RCTs in educational research include concerns about denying students access to treatment, the difficulty of standardizing interventions in school settings, and low external validity (Asmussen, 2011; Cartwright & Hardy, 2012; Hanley, Chambers, & Haslam, 2016). Wrigley (2016) highlights four factors that make it difficult to transfer RCTs to school settings: Children in classes cannot be randomly allocated; it is difficult for teachers to alter practice; as previously mentioned, there are ethical issues in non-treatment of control groups; and experimental methods require that all other factors be frozen and "children don't freeze easily" (p. 239). Each of these factors limits the external validity, or generalizability, of the research results. While RCTs may control for many threats to internal validity, they may not be practical or relevant to education research. Furthermore, the validity and utility of other research methodologies may be underestimated due to the emphasis on RCTs (Kourea & Lo, 2016).

Single-case design. Single Case Design studies are, "experimental methods consisting of various designs involving repeated measures of a specific behavior or skill under different conditions to evaluate the effectiveness of a treatment for an individual or a small group of

individuals that serve as their own control" (Hitchcock, Kratchowill, and Chezan, 2015, p. 460). This type of research is considered to be "a rigorous, scientific methodology used to define basic principles of behavior and establish evidence-based practices" (Horner et. al., 2005, p. 165). It is ideally suited to demonstrating intervention outcomes in school settings since it does not have the methodological challenges inherent in RCTs (Burns, 2012).

While SCDs demonstrate strong internal validity, the results "are not likely to capture patterns in the data across time, could miss idiosyncrasies in the data, and are overly affected by atypical baseline data" (Burns, 2012, p. 177), thus affecting external validity. The What Works Clearinghouse, however, accepts SCDs as "evidence of causal validity" (Burns, 2012, p. 177) and provides standards for evaluating them, including internal and external validity criteria (Hitchcock et al., 2015).

Best available evidence. Questions regarding what is considered to be the "best" research evidence in education have contributed to the research-to-practice gap. However, it is critical to note that the evidence-based practice model refers to the best *available* evidence, rather than the *best* evidence. According to Spencer et al. (2012), "the term *best available evidence* implies that there is a range of evidence and that educators should select the best of what is available" (p. 133). In this sense, the RCT versus SCD debate can be resolved on a case-by-case basis, depending upon the question and contextual variables that practitioners are attempting to address.

While the term *best available evidence* is often considered to be synonymous with high standards of research methodology (i.e. RCTs), that perspective may "limit the scope and impact of evidence-based practice to those educational decisions on which high quality intervention research is plentiful and definitive leaving educators with little guidance on the majority of the

decisions they face" (Detrich et al., 2012, p. 153). In other words, high quality research studies do not exist to address every problem that educators face; thus, they must rely on what they do have access to. Additionally, practitioners tend to primarily be concerned with interventions and practices that are relevant to their particular context. In this scope, the effectiveness of the intervention may be outweighed by the availability of the intervention.

Sustaining empirically supported practices. While many researchers have investigated educators' views of research, few have examined the effects of teachers' use of research in the classroom. Huat See et al. (2016) conducted a study of teachers' engagement with research evidence through the use of enhanced feedback. Nine schools in England participated in the study, in which the schools were responsible for devising and conducting the intervention, which was evaluated, observed, and monitored by the researchers. They found that, while teachers were willing and able to engage with the research, they only implemented the empirically supported intervention at a surface level, indicating that they did not fully comprehend the research.

Additionally, the study found that the teachers did not consistently implement the intervention with fidelity. Though the teachers had been cautioned against using feedback directed at the self, rather than toward self-regulation, they continued to do so (Huat See et al., 2016). In addition to, or perhaps because of the teachers' difficulty with implementing the intervention, the researcher found the intervention to have minimal impact on student outcomes. This suggests a need for extensive professional development and training to fidelity to increase sustainability of empirically supported practices (Baker Gersten Dimino, & Griffiths, 2004; Williams, Lee, Harrison, & Black, 2004). It is evident, that educators require high levels of support to effectively implement research-based interventions.

Measuring EBP Use in Schools

A paucity of instruments to assess factors related to EBP exists. The available tools have been developed primarily in the fields of medicine and mental health. Even so, few instruments refer specifically to EBP, targeting instead readiness to implement organizational change or innovations and intending to assess implementation of a specific intervention or introduction of a particular innovation.

Education. In the field of education, specific instruments to assess factors related to EBP do not currently exist. The primary instrument cited in research studies is the *Innovation Configuration Component Map* (ICCM), which is an observation tool that "assumes that individuals implement any innovation in different ways and with varying levels of quality" (Beauchat, Blamey, & Walpole, 2009, p. 27). The term *innovation* refers to interventions, programs, or instructional strategies. The ICCM "identifies essential components of the practice and provides concrete examples from ideal implementation to least effective" (p. 27). To utilize an ICCM, an organization must have a specific intervention in mind for observation, as it cannot be applied generally. For example, Beauchat et al. (2009) developed an ICCM for shared storybook reading with preschool students. Javeri and Persichitte (2007) developed an ICCM to measure technology innovation practices among higher education faculty, while, Craig and Kacer (2000) created an ICCM to assess the relationship between student achievement and implementation of extended school services.

The ICCM provides guidelines for researchers and practitioners to follow in order to develop an assessment specific to their particular innovation. An ICCM is composed of *components*, *variations,* and *configurations. Components* refer to major features of the innovation; *variations* are different ways in which the components can be operationalized; and *configurations* are the operational patterns that result from the selection and use of different

component variations (Javeri & Persichitte 2007). For example, in developing an ICCM for shared storybook reading, one *component* is "modeling rich language". There are four variations of the component, ranging from *ideal* to *least effective*. In the *ideal* variation "the teacher intersperses descriptive language while reading and explaining story content", whereas in the *least effective* variation "the teacher reads the words in the story and does not add any language" (Beauchat et al., 2009, p. 29). In this case, a *configuration* would refer to the use of different variations of each of the shared storybook reading components.

There are several disadvantages of the ICCM, including that each innovation requires the development of a new instrument. Additionally, the purpose of the instrument is to assess implementation of an empirically supported intervention, rather than implementation of the broader definition of evidence-based practice. An observer completes the survey while watching the intervention being implemented. The ICCM does not address practitioner involvement in the selection and implementation of the intervention, nor does it provide quantitative data related to EBP attitudes, barriers, or implementation.

Healthcare. Like in the field of education, in the medical field instruments to assess facets of evidence-based practice are scarce. Most instruments are intended to assess organizational readiness to implement a particular intervention. Furthermore, few instruments specifically reference *evidence-based practice,* referring primarily to organizational *change.*

Organization Readiness to Change Assessment (ORCA). The *Organization Readiness to Change Assessment* (ORCA) was developed for use in quality improvement projects to assess overall organizational readiness and to identify specific barriers or challenges to innovation implementation (Helfrich, Li, Sharp, & Sales, 2009). While the instrument examines general

organization readiness to implement change, the organization must, again, have a specific program or intervention in mind.

The survey is composed of three major scales: (1) Strength and extent of evidence for the clinical practice changes (Evidence), (2) Quality of organizational context (Context), and (3) Capacity for internal facilitation (Facilitation). Within the Evidence scale are three subscales: discord (disagreements among employees/ team members about evidence), research evidence, clinical experience, and patient preferences. The Context scale consists of six subscales: two subscales to assess organizational culture, two subscales assess leadership practice, one subscale to assess evaluation in terms of setting goals, and one to assess resources for supporting organizational changes. Finally, the Facilitation scale includes nine elements related to the organization as a whole: (1) characteristics of senior leadership, (2) characteristics of "clinical champions", (3) opinion leader roles, (4) implementation team member roles, (5) implementation plan, (6) communication, (7) implementation progress, (8) implementation resources, and (9) implementation evaluation (Helfrich et al., 2009). Seventy-seven Likert-type items comprise the scale, with ratings from 1 (strongly disagree) to 5 (strongly agree).

In a study examining the psychometric properties of the ORCA, Helfrich et al. (2009) administered the survey to 80 staff participating in three quality improvement projects: the Cardiac Care Initiative, the Lipids Clinical Reminders project, and an intensive care unit quality improvement project. Results of reliability analysis of the three scales and constituent subscales found mediocre reliability for the *Evidence* scale and its subscale, which could be attributed to too few items and a small sample size. Additionally, results of factor analysis indicated that two subscales, one from the Facilitation scale and one from the Context scale, failed to load on any of the three factors. The study also failed to address validity of the instrument as a predictor of

evidence-based clinical practice. It is evident that additional research should be conducted to further investigate the psychometric properties of the ORCA.

***Organizational Readiness for Implementing Change* (ORIC)**. This assessment was developed for a similar purpose to that of the ORCA. Organizational readiness, in relation to the instrument, refers to "the extent to which organizational members are psychologically and behaviorally prepared to implement organizational change" (Shea, Jacobs, Esserman, Bruce, & Weiner, 2014). Whereas the ORCA concentrated on assessment of the individual, the ORIC focuses on supra individual (team, department, or organization) level assessment. This instrument includes two facets of readiness to implement change: change commitment and change efficacy. Four studies were conducted to examine the psychometric properties of the ORIC.

The purpose of the first study was to determine content adequacy, or the degree to which the survey items represented the theoretical construct. The second study ascertained factor structure and reliability of the instrument at the individual level. Study three was intended to ascertain reliability and validity of the instrument at the organizational level. Finally, study four was a pilot study of the instrument with organization employees, rather than university students posing as employees, as were the previous studies. The final study employed confirmatory factor analysis to confirm the previously identified factor structure of the instrument (Shea, et al., 2014).

Results of these studies indicated that the ten items included in the ORIC reflected the theoretical content of change commitment and change efficacy and that the instrument consisted of two correlated factors with good item fit and high factor loadings. Reliability analysis yielded high inter-item consistency, as well as inter-rater reliability and inter-rater agreement that

supported organizational level analysis (Shea et al., 2014). Similar to the ICCM and the ORCA, the ORIC is designed to assess organizational readiness to implement a particular change. The items are worded in such a way, that the respondents must have a specific change in mind; for example, "we are committed to implementing *this change*."

The ORIC and the ORCA are the primary instruments cited in the literature used in healthcare to assess organizational change. While their psychometric properties have been examined, certain questions remain. The validity of the ORCA, for example, has yet to be verified. There are also concerns related to scale reliability and the factor loadings of two of its subscales. The ORIC, however, demonstrates good reliability and a confirmed factor structure, but only consists of ten items. The primary disadvantage of the two instruments is that they are designed to assess readiness to implement specific changes or innovations.

The Evidence-Based Practice Attitude and Utilization Survey (EBPASE). Unlike the ICCM, the ORCA, and the ORIC, the EBPASE is an instrument intended to assess chiropractors' general attitudes, barriers, and use of evidence-based practice. The 84-item self-report survey is composed of three subscales: attitudes, skills, and use of EBP, with the other sections capturing "facilitators and barriers of EBP uptake and EBP training, and participant demographic characteristics " (Terhorst, Leach, Bussieres, Evans, & Schneider, 2016, p. 328). Items are rated on a 5-point Likert Scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The instrument has demonstrated good internal consistency reliability (Cronbach's alpha= 0.84) and content validity (CVI=0.90). It has also demonstrated moderate test-retest reliability (ICC = 0.578–0.986) (Bussieres, Terhorst, & Leach, 2015).

While the EBPASE is intended to assess general attitudes, barriers, and use of EBP, the questions are intended for health professionals. The "awareness of clinical practice" subscale

references guidelines that are specific to the medical field. Additionally, the scale includes 84 items, 14 of which are demographic items intended for chiropractors. The scale consists of seven "parts" (Part A through Part F), which employ 6 different Likert-type response scales, ranging between four and six points. The length of the scale, its intended population, and the variations in response type are potential disadvantages of employing the survey with educators.

Mental health. In the mental health field, few assessment instruments are being studied in relation to organizational change and innovation implementation. The Texas Christian University (TCU) *Organizational Readiness for Change* (ORC) instrument has been cited in several research studies. The ORC was developed to "represent the most relevant variables for studying innovation and change efforts in substance abuse agencies" (Lehman, Greener, & Simpson, 2002, p. 197). There are different versions of the ORC, which are dependent upon the identity of the respondents. For example, the ORC-D is intended for organization directors to complete while the ORC- S is intended for organization staff.

The ORC includes 115 Likert-type items, which represent 18 content domains. The domains include multiple scales in four major areas: motivation for change, institutional resources of the program, personality attributes of the staff, and organizational climate of the program (Lehman et al., 2002). Reliability analyses yielded reasonably acceptable reliability coefficients. Additionally, results of the analyses indicated that the eighteen scales are "generally unidimensional" (p.207). In contrast with the ICCM, the ORCA, and the ORIC, the ORC is designed to broadly assess an organization's readiness for implementing change, as opposed to targeting a single innovation or program change. It can also differentiate between attitudes and perceptions of program staff and program directors.

While the ORC has some significant advantages, including promising reliability coefficients, the number of items is a potential disadvantage in employing the instrument with teachers, administrators, and related service personnel. Additionally, the items are intended towards individuals in mental health settings. Since the survey has two variations: one for staff and one for program directors (i.e. administrators), it would require that both surveys be administered within the setting. Finally, the complexity of the instrument may act as a barrier in applying it in education.

Feasibility, acceptability, and usability. Each of the existing assessment instruments has strengths, including the ability to assess multiple dimensions of EBP or readiness for organizational change and the ability to obtain multiple perspectives. However, no single instrument encompasses attitudes, barriers, and actual implementation of EBP in school settings. Additionally, most of the instruments consist of 50 to 100 items. Research indicates that one of the primary reasons that teachers do not engage in research-related activities is a lack of time (Cain, 2016; Gore & Gitlin, 2004). Thus, survey instruments must be brief. Educators also prefer to engage in research activities (i.e. completing a survey) when they feel that the activity is practically relevant to their classroom experience (Gore & Gitlin, 2004). So, it is essential that the survey's purpose be evident to those completing it. In this case, the objective in assessing teachers', administrators', and related service providers' attitudes toward and use of EBP would be to inform professional development activities.

Instrument Development

As the development of assessment instruments is a complex and difficult process, Gall, Gall, and Borg (2007) suggest that researchers "make certain that no suitable test is available" (p.222) before researchers consider developing a new instrument. A review of the literature

revealed that no single existing instrument encompasses attitudes, barriers, and use of EBP in school settings. The measures that are currently available tend to focus on specific treatments or interventions, rather than the broad construct of EBP. The development of a brief, comprehensive measure of evidence-based practice use could inform school leaders' and administrators' professional development considerations and facilitate the use of EBP in schools. However, there are several considerations that should be made in the development of new survey instruments, including: item selection, reliability, validity, and practicality. Each of which will be discussed further in the following sections.

Item selection and scoring. Gall et al. (2007) identified seven steps in survey construction. The first is to define the construct to be measured, followed by defining the target population, and reviewing related tests. All of which, are instrumental in identifying or developing test items. Some considerations in developing test items include: the inclusion of reverse items, item length, and item scoring (Hamby, Icke, & Babcock, 2016). Many researchers advocate for the inclusion of reverse scored items, arguing that they act as mental "speed bumps", causing respondents to stop and think before responding. Reverse items may also cause respondents to think of the construct from different perspectives (Hamby et al., 2016). However, the inclusion of reverse items may result in lower item-total correlations and lower scale reliability. Additionally, reverse items have a tendency to complicate an instrument's factor structure, resulting in a separate factor for reverse-scored items.

Similar to the inclusion of reverse-scored items, item length can impact scale reliability and validity. There are a variety of reasons to keep test items brief, including a decreased likelihood of misinterpretation and increased comprehension. Additionally, average item length has been found to be negatively associated with scale reliability (i.e. Cronbach's alpha) (Hamby

et al., 2016). Studies have also found average number of words to be negatively associated with the proportion of respondents selecting the same response for the same item in two administrations of the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1940).

In regards to survey scoring, Likert scales are one of the most popular tools for measuring psychological constructs. However, the best method of structuring them is still a matter of debate (Maeda, 2015). The general structure consists of item stems (phrases or statements) and scales (response options), which are rated along a continuum (i.e. strongly disagree to strongly agree). Some studies suggest that the way in which the responses are arranged (i.e. positive on the left/negative on the right or vice versa) can affect the way participants respond. For example, Hartley and Betts (2010) found that a test with items that started with a positive label and had the highest numerical rating on the left produced significantly higher results than three other versions of the same test. Maeda (2015) suggests utilizing vertically oriented response options, rather than horizontally oriented response options to negate left-side selection bias.

The optimal number of response options is also a matter of debate, with responses ranging from 2 to 100 options. Though research suggests that surveys with 4 to 7 response options yield the highest reliability coefficients (Nadler, Weston, & Voyles, 2015). When determining the number of response options, some researchers have argued against inclusion of a neutral midpoint (i.e. neither agree nor disagree), contending that participants are more likely to select a neutral option if one is offered (Nadler et al., 2015). Exclusion of the neutral midpoint has also been shown to reduce social desirability bias (Garland, 1991). Finally, Nadler et al. (2015) found that respondents have different interpretations of the midpoint, including: “no opinion”, “unsure, and “neutral”. These findings indicate that researchers’ and participants’

perception of the midpoint may differ, thus impacting the study results and potentially, the reliability of the data.

Development of the EBPSAT. The objective in designing the EBPSAT was to assess attitudes, barriers, and use of EBP in school settings. After an extensive literature review, the researchers identified three primary areas of focus: EBP at the organizational level (i.e. support and/or barriers at the school/district/clinic level), EBP at the individual level (i.e. personal attitudes toward EBP), and practice of EBP (i.e. specific examples of EBP use in practice). The second step in the process was to generate a list of potential survey items. The original list consisted of 53 items. The researchers selected twenty items for inclusion in the instrument. Five items comprised the *Organization* subscale, eleven were included in the *Individual* subscale, and the *Practice* subscale consisted of four items. It was essential that the instrument be brief, comprehensive, and easy to understand.

The third step in developing the EPBSAT was to determine how items would be rated. Likert scales are one of the most popular tools for measuring psychological constructs. According to Maeda (2015), the best method of structuring Likert scales is still under debate. The general structure consists of item stems (phrases or statements) and scales (response options), which are rated along a continuum (i.e. strongly disagree to strongly agree). A four point Likert scale was selected for the EBPSAT with ratings from 1= strongly disagree to 4= strongly agree. The *Organization* subscale also included a neutral option, 0= I don't know. Finally, the four items on the *Practice* subscale were rated dichotomously (1=yes, 0= no).

Reliability. Test reliability refers to "the degree to which measurement error is absent from the scores yielded by the test" (Gall et al., 2007, p.200), where "measurement error" is defined as the difference between an individual's *true* score and his *obtained* score on a test.

Since true score and measurement error are both hypothetical constructs, they cannot be measured directly. There are, however, several methods of obtaining reliability estimates by calculating reliability coefficients, whose values ranging from 0.0 to 1.0, with a score of 1.0 indicating perfect reliability and a score of 0.0 indicating zero reliability (Gall et al., 2007). Generally, reliability coefficients of 0.80 or better are considered to be sufficiently reliable for research purposes. According to Gall et al. (2007), reliability is essential to validity. However, having good reliability does not automatically imply validity.

There are several approaches to estimating test score reliability, including: alternate form, test-retest, and internal consistency. Alternate forms reliability involves the development of different versions of the same test and correlating participants' scores on each test. Whereas, test-retest reliability requires that the participants complete the same test on two separate occasions and correlating participants' scores on each administration. Internal consistency reliability is particularly relevant to the current study; it involves examining the individual items on the test and is described further in the following section (Gall et al., 2007).

Internal consistency. According to Dunn, Baguley, and Brunsden (2013), internal consistency estimates were the most common reliability estimate reported in articles published by the American Psychological Association (APA). There are several methods for estimating internal consistency reliability. Each of which, involves analyzing a sample of scores from a single administration of the test. For example, a split-half correlation coefficient can be calculated by splitting the test into two subtests (usually by placing odd numbered items on one subtest and even numbered items on the other). However, the resulting reliability coefficient only represents half the test, thus the Spearman-Brown prophecy formula is utilized to correct the reliability coefficient to account for the entire test (Gall et al., 2007).

Another method of approaching internal consistency reliability is the method of rational equivalence, which involves analyzing the individual test items using one of several formulas (Gall et al., 2007). The most common formulas include the Kuder-Richardson formulas (K-R 20 and K-R 21) and Cronbach's alpha coefficient. In order to utilize the Kuder-Richardson formulas, items must be scored dichotomously. Both K-R 20 and K-R 21 yield lower reliability coefficients than would be obtained by utilizing other means. Cronbach's alpha coefficient, or Cronbach's alpha, is "a general form of the K-R 20 formula that can be used when items on a measure are not scored dichotomously" (Gall et al., 2007, p. 202). It is the most widely reported estimate of internal consistency reliability (Furr & Bacharach, 2014).

Validity. The 1999 *Standards for Educational and Psychological Testing* defined validity as the, "degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (American Educational Research Association, 1999, p. 9). It is important to highlight the definition's emphasis on *interpretation* of test scores, rather than the test scores themselves, as test scores are not inherently valid or invalid (Gall et al., 2007). In the early years of test development, researchers referred to different *types* of validity (content, construct, correlational, criterion, concurrent, and predictive). However, Messick (1989) proposed a unified model in which, all validity is of a single type supported by multiple sources of evidence. Thus construct validity became content evidence, criterion validity became criterion-related evidence, and so on. This change was more than simply semantic. By thinking about validity *evidence*, rather than validity *types*, researchers began to view validity as hypotheses about score interpretation and, thus subject to hypothesis testing (Cook, 2014). Three sources of validity evidence are discussed further in the following sections: content validity evidence, construct validity evidence, and criterion-related evidence.

Content validity evidence. Content validity evidence demonstrates, "the extent to which the items on a measure assess the same content or how well the content material was sampled in the measure" (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003, p. 94). Content validity evidence can be characterized as face validity evidence, which refers to a measure appearing to be valid at face value, and logical validity evidence, which requires more rigorous methods of analysis. Polit and Beck (2006) describe two distinct phases of judgment in evaluating logical validity evidence: "a priori efforts by the scale developer to enhance content validity through careful conceptualization and domain analysis prior to item generation, and a posteriori efforts to evaluate the relevance of the scale's content through expert assessment" (p. 490).

Content validity index. There is a multitude of methods of quantifying expert agreement regarding the relevance of survey items, including: averaging expert ratings using pre-established criteria, utilizing coefficient alpha, and computing multi-rater kappa coefficient (Polit & Beck, 2006). However the content validity index (CVI) is the most widely reported measure of content validity evidence in the literature (Newman, Lim, & Pineda, 2013). The CVI involves assembling a panel of experts in the content area; having them indicate whether each item on the scale is relevant to the construct; computing the percentage of relevant items for each expert; and taking an average of the percentages across experts (Rubio et al., 2003).

While the CVI is widely reported in the literature, it does have its critics. Some argue that the CVI does not adjust for chance agreement and should be supplemented by (or substituted with) the kappa coefficient, which "yields an index of degree of agreement beyond the chance agreement" (Polit & Beck, 2006, p. 490). Others point out that important information is lost when the CVI collapses multipoint ratings into two categories (relevant/not relevant). The CVI also fails to capture the comprehensiveness of the item set.

Construct validity evidence. Construct validity evidence refers to evidence of "the degree to which an instrument measures the trait or theoretical construct that it is intended to measure" (Bolarinwa, 2015, p. 197). Bolarinwa (2015) describes construct validity as the "most valuable and most difficult measure of validity" (p. 197). There are four types of construct validity evidence: convergent validity, discriminant validity, factorial validity, and known groups validity. Both factorial validity and known groups validity are of particular importance in the current study and are discussed further below.

Factorial validity. Factorial validity is an "empirical extension" (Bolarinwa, 2015, p. 198) of content validity, which is established through the statistical technique of factor analysis (FA). FA evaluates the underlying factor structure of the construct the instrument is intended to measure by reducing a large number of items into a small number of factors. Cronbach and Meehl (1955) described the use of factor analysis to provide evidence of construct validity thus,

> If a trait such as dominance is hypothesized, and the items inquire about behaviors subsumed under this label, then the hypothesis appears to require that these items be generally intercorrelated. Even low correlations, if consistent, would support the argument that people may be fruitfully described in terms of a generalized tendency to dominate or not dominate (p. 288)

Essentially, the statistical analysis provides evidence (or lack of evidence) of the existence of the hypothesized construct.

Known groups validity. Known group validity evidence is established when "a group with already established attribute of the outcome of construct is compared with a group in whom the attribute is not yet established" (Bolarinwa, 2015, p. 197). Cronbach and Meehl (1955) stated it simply, "If our understanding of a construct leads us to expect two groups to differ on

the test, this expectation may be tested directly" (p. 287). For example, if an instrument were designed to measure depression, known groups validity could be established by administering the assessment to a sample of individuals diagnosed with depression and comparing scores from that sample with those of a sample of individuals from a different population, the hypothesis being that those diagnosed with depression would score higher on the measure. If the test is able to discriminate between groups, there is evidence of its usefulness as a decision-making instrument (Hattie & Cooksey, 1984).

Criterion-related evidence. The final source of validity evidence, criterion-related evidence, is a measure of how well a test compares to another measure of the same construct (Bolarinwa, 2015). It includes two subtypes: concurrent validity, which measures the new test against a "highly rated existing standard (gold standard)" (Bolarinwa, 2015, p. 197) and predictive validity, which evaluates the instrument's ability to "forecast future events" (p. 197). A limitation of criterion-related evidence is that it is dependent upon having a criterion, making it difficult to establish for newly hypothesized constructs. If a criterion measure does exist, criterion-related evidence is dependent upon the quality of the criterion variable. Finally, a correlation between criterion measures does not equate to equivalence. If the measure correlates with multiple variables, it may be difficult to determine what is actually being measured.

Professional Development Considerations for School Leaders

Professional development (PD) is a process, "whereby an individual acquires or enhances the skills, knowledge, and/or attitude for improved practice" (Mitchell, 2013, p. 390). Like the use of EBP, professional development is a good idea in theory, but there are some challenges in its execution. For example, PD generally takes place in a single day, which yields minimal improvement in teacher performance (Desimone, Porter, Garet, Yoon, & Birman, 2002).

Teachers are also given little opportunity to practice their newly learned skills and receive little feedback on their performance (Yoon, Duncan, Lee, Scarlos, & Shapley, 2007). Furthermore, there is little evidence to indicate that school districts' decisions about which professional development opportunities to offer are based on quantifiable data. With the changing educational landscape, school leaders should focus on more personalized learning opportunities for teachers, taking into consideration the influence of perceptions, attitudes, and philosophies in addition to skills deficits (Stevenson, Hedberg, O'Sullivan, & Howe, 2016).

Brownlee et al. (2015) call for an "evaluative stance" (p. 412), defined as "a mindset and skillset focused on making evidence informed decisions" (p.412) on in decision-making about professional development. In interviewing six Australian school leaders' to examine their approaches to decision-making related to PD, they identified three categories of influence: intrapersonal influences, leadership style and skills, and contextual factors. Intrapersonal influences included leaders' sense of self-efficacy, their professional identity, and beliefs about professional practice. Leadership style and skills refer to a school leaders' ability to work collaboratively with others to identify and prioritize goals. Finally, contextual factors include organization vision, priorities, and budget (Brownlee et al., 2015). Two of the six participants directly referenced factors relating to an "evaluative stance", while the other four participants made comments alluding to an "evaluative mindset" (p.425) in PD decision-making.

In a study of school leaders' approaches to PD in response to a curriculum change, Stevenson et al., (2016) found that one of three Australian schools in the study conducted needs-based assessment to inform PD activities, while the other two schools assumed areas for growth and developed PD activities without teacher input. The results indicated that the largest impact occurred in the school in which leaders asked all teachers to identify professional needs and then

formalized approaches to address them; whereas, the other two schools observed pockets of success among specific teachers or teams (Stevenson et al., 2016).

Similarly, Castillo, March, Yin Tan, Stocklager, Brundage (2016), assessed educators perceptions of response-to-intervention (RtI) before employing targeted PD activities with the aim of changing teacher perceptions, in addition to increasing skill; their rationale being that numerous studies indicate that it is important to target teachers' beliefs when attempting to change behavior. The study took place over the course of three years. During which, they found, that among a sample of 3,961 educators across a southeastern state, membership on School Leadership Teams (SBLT) predicted positive changes in perceptions of RtI. However, direct, intensive training in RtI did not have a statistically significant correlation to changes in belief (Castillo et al., 2016). These results indicate that it is essential to use data derived from teacher reports of their needs to inform PD. Especially since several studies have indicated that the more influence teachers have in tailoring PD activities, the more likely they are to find it effective (Leask & Younie, 2013).

Summary

Legislation mandates the use of research evidence in education to improve student outcomes, but there is little evidence that teachers are doing it. The existence of the research-to-practice gap indicates that there are significant barriers between knowledge generation and its use in the field. Such barriers include: teachers' attitudes toward research, challenges in dissemination, and questions about what constitutes "evidence". Additionally, there are few instruments for assessing the degree to which teachers and related service providers are utilizing EBP in their daily practice. The measures that are currently available tend to focus on specific treatments or interventions, rather than the broad construct of EBP. The development of a brief,

comprehensive measure of evidence-based practice use could inform school leaders' and administrators' professional development considerations and facilitate the use of EBP in schools.

The EBPSAT is a single page survey, specifically designed for use in school settings. Consisting of only 20 items, the survey can be completed in less than ten minutes to accommodate educators' busy schedules. The instrument's three subscales are intended to quickly assess attitudes, barriers, and use of EBP to identify areas for growth to inform district's professional development decisions.

Chapter 3

Method

The following sections describe the method employed in this study. First, an overview of the research design will be discussed, followed by a discussion of the selection of participants. This will be followed by a description of the instrumentation used, as well as a description of the data collection procedures. Finally, this section concludes with a description of the statistical analyses that were employed to answer each research question.

Research Design

This study is an exploratory study designed to examine the psychometric properties of the newly developed EBPSAT (Spencer, 2016). A multivariate correlational design was used to examine questions of reliability, validity, and the underlying factor structure of the EBPSAT. Multivariate statistics allow for the analysis of multiple independent variables (IVs) and/or dependent variables (DVs) by correlating the variables with one another. These methods allow researchers to examine complex research questions (Tabachnick & Fidell, 2013). Specifically exploratory factor analysis (EFA) and analysis of variance (ANOVA) were employed.

Correlational research designs can be used for either of two purposes: to determine the degree to which a relationship exists between two or more variables or to develop prediction models to be able to predict the future value of a variable from the current value of one or more other variables. For the purpose of the current study, a correlational design was used to determine the degree to which relationships exist between variables such as individual survey items and the underlying factors of the EBPSAT.

A survey research approach was utilized to obtain the necessary data to conduct multivariate correlations. Survey research is a type of descriptive research, in which data are

collected via questionnaires or interviews regarding participants' beliefs, attitudes, interests, and/or behavior. This information can be gathered through the use of paper and pencil tests or through web-based survey software. For the purpose of the current study, web-based survey software (Survey Monkey) was utilized.

Selection of Participants

Sample. The target population for this research study includes teachers, administrators, and school-based related service providers in the United States. However, the accessible population is restricted to school personnel who can be contacted via social media, educator message boards, and Educator listservs. A convenience sample, composed of participants who respond to an electronic survey link was employed. Institutional Review Board (IRB) approval from Northern Arizona University was obtained prior to data collection (see Appendix A). As this study posed minimum risk to participants and did not target special or vulnerable populations, it was awarded exempt status by the IRB.

While there is a significant amount of debate over the adequate sample size necessary for conducting exploratory factor analysis, there is no consensus on a set guideline. MacCallum, Widaman, Zhang and Hong (2001) and Thompson (2000) recommends sample sizes between 100 and 200. A sample size of at least 300 is recommended when communalities are low, there are a small number of factors, or only three or four indicators are present for each factor (MacCallum et al., 2001). The recommended ratio of participants to measured variables, according to Thompson (2000) is 10:1. However, Bentler and Chou (1987) suggest a ratio of 5:1. For the purpose of this study, a sample size between 300 and 500 was attempted. The final sample size of 291 (with missing data and outliers deleted) resulted in a ratio of 15:1. As the

factor loadings were moderate to high, the sample is considered to be adequate (291:20 > 5:1). The demographic information for the sample is outlined in Table 1.

Table 1

Participant Demographics

	Frequency	Percentage
Gender		
Female	226	77.7
Male	63	21.6
Other	1	.3
Race/Ethnicity		
White	240	82.5
Hispanic	23	7.9
Black/African American	11	3.8
Other	11	3.8
Asian/Pacific Islander	5	1.7
Primary Language		
English	289	99.3
Spanish	1	.7
Age Group		
18-24	4	1.4
25-34	60	20.6
35-44	77	26.5
45-54	84	28.9
55+	65	22.3
Highest Degree		
HS Diploma or Equivalent	1	.3
Some College, No Degree	1	.3
Associate's Degree	1	.3
Bachelor's Degree	19	6.5
Master's Degree	163	56
Professional Degree	51	17.5
Doctoral Degree	54	18.6
Research Courses		
Yes	240	82.5
No	50	17.5
Title/Position		
Paraprofessional	1	.3
Teacher	54	18.6
Administrator	131	45

	Frequency	Percentage
Intervention Specialist	1	.3
Speech Language Pathologist	4	1.4
School Psychologist	55	18.9
Occupational Therapist	1	.3
Other	42	14.5
Years at Current Position		
Less than 1 year	37	12.8
1-5 years	135	46.4
6-10 years	42	14.4
11-15 years	34	11.7
16-20	26	8.9
More than 20 years	16	5.5
Setting		
School	268	92.1
Clinic/Hospital	2	.7
Other	20	6.9

Note. N=290. Values were rounded, therefore may not add up to 100%.

Of the 291 participants, 77% (n=226) were female, 21.6% were male (n=63), and 0.3% (n=1) selected "other". The majority of participants identified as white (n=240, 82.5%), with 7.9% Hispanic (n=23), 3.8% Black/African American (n=11), 3.8% other (n=11), and 1.7% Asian/Pacific Islander (n=5). All but one participant (n= 289, 99.3%) identified their primary language as English, with a single participant (0.7%) indicating Spanish as their primary language. The age of participants was relatively evenly distributed with the exception of the 18-24 year age range, which comprised 1.4% (n=4) of the sample. Individuals in the 25-34 year age range (n=60) comprised 20.6% of the sample, those in the 35-44 year range (n=77) 26.5%, individuals 44-54 (n=84) 28.9%, and individuals age 55 and older (n= 65) comprised 22.3% of the sample.

The majority of participants in the sample (n=163, 56%) reported that they held a master's degree, while 18.6% (n= 54) held doctoral degrees, 17.5% (n=51) held professional degrees, and 6.5% (n=19) held bachelor's degrees. One participant (0.3%) reported holding an

associate's degree; one (n=0.3%) reported some college without a degree; and one (n=0.3%) reported having a high school diploma or equivalent. Of the 290 participants, 82.5% (n=240) indicated that they had taken research courses in college. Regarding the length of time participants had spent at their current position, 46.4% (n=135) reported1-5 years; 14.4% (n= 42) reported 6-10 years; 12.8% (n=37) had been at their position for less than one year; 11.7% (n=34) reported 11-15 years; 8.9% (n=26) reported 16-20 years; and 5.5% (n=16) reported having been at their position for more than twenty years.

The majority of participants (n=262, 92.1%) reported working in a school setting, compared to two participants (0.7%) working in clinic or hospital setting, and 6.9% (n=20) working in a setting other than a school or clinic. Nearly half of the participants (n=131, 45%) reported their position as administrators, with 18.9% (n=55) being school psychologists, and 18.6% (n=54) being teachers. It should be noted that, though 54% of the sample reported being teachers, when asked to identify the grades and subjects taught, 23.4% (n=68) responded. Teacher demographics are illustrated in Table 2.

Table 2

Teacher Demographics

	Frequency	Percentage
Grade		
Kindergarten	1	1.5
First Grade	1	1.5
Second Grade	2	2.9
Third Grade	1	1.5
Fourth Grade	2	2.9
Fifth Grade	1	1.5
Middle School (6-8)	14	20.6
High School (9-12)	36	52.9
More than 1 grade	10	14.7
Subject		

	Frequency	Percentage
English/Language Arts	7	10.3
Math	6	8.8
Science	18	26.5
History	4	5.9
Music	10	14.7
Multiple Subjects	18	26.5

Note. N=68. Values were rounded, therefore may not add up to 100%.

Of the 68 teachers in the sample, 52.9% (n=36) were high school teachers; with 20.6% (n=14) being middle school teachers (grades 6-8); 14.7% (n=10) reported teaching more than one grade; 2.9% (n=2) reported teaching each second and fourth grade; and 1.5% (n=1) reported teaching each kindergarten, first grade, third grade, and fifth grade. Science teachers comprised the largest percentage of the sample with 26.5% (n=18), while 26.5% (n=18) reported teaching more than one subject. Music teachers comprised the next largest portion of the sample at 14.7% (n=10), with English/Language Arts teachers comprising 10.3% (n=7) of the sample, math teachers 8.8% (n=6), and history teachers 5.9% (n=4).

Expert Panel. Twenty-two professionals who have expertise in evidence-based practice were identified. Inclusion criteria for the expert panel included having an advanced degree (doctorate or master's degree), having conducted extensive research in the area of EBP, and having publications on the subject of EBP in peer reviewed journals, as well as having presented evidence-based topics at conference presentations. Of those, twelve completed the content validity questionnaire. Of the twelve experts, eight were male (67%) and four female (33%). Eleven experts held doctorates and one held a master's degree. The panel included five professors, one associate professor, one assistant professor, and five who reported their position as "other". Those that marked "other" as their position included: one researcher, one senior fellow, one director, and two professor/division chairs. All twelve panelists reported that they

had published evidence-based practice articles in peer-reviewed journals and presented evidence-based topics at professional conferences. Table 3 contains expert demographic information.

Table 3

Expert Demographics

	Frequency	Percentage
Gender		
Male	8	67
Female	4	33
Education Level		
Doctoral Degree	11	92
Master's Degree	1	8
Title/Position		
Professor	5	42
Associate Professor	1	8
Assistant Professor	1	8
Other	5	42
Published EBP Articles		
Yes	12	100
No	0	0
Presented EBP Topics		
Yes	12	100
No	0	0

Note. N=12

Instrumentation

Evidence Based Practice Self-Assessment Tool. The EBPSAT (see Appendix F) is a 20-item survey, which is composed of three subscales: *Organization, Individual,* and *Practice*. The first two subscales, *Organization* and *Individual*, include 15 Likert-type items with responses ranging from 1 (strongly agree) to 4 (strongly disagree). The *Organization* subscale also includes a 0 response indicating "I don't know". The inclusion of a *Don't Know* response was necessitated by the fact some individuals may be unaware of the organization's implementation of EBP. Though the inclusion of a *Don't Know* response option decreased the

number of responses used in the data analysis, some research has shown that a *Don't Know* option can increase the reliability of responses by decreasing random responding (Dedrick & Greehbaum, 2011). Analysis of the *Don't Know* responses is provided in the results. The *Organization* items are intended to assess an individual's perception of EBP implementation at the school or district level. While, the *Individual* subscales examines personal perceptions and values of EBP facets. Finally, the *Practice* subscale includes five dichotomous items (i.e. 0=no, 1=yes) regarding EBP activities in which the respondent has engaged in the past eight weeks. The survey yields a total score (0-68).

EBP Self-Assessment Tool Content Validity Questionnaire. A content validity questionnaire was developed to assess the relevance of each of the items included in the EBPSAT (see Appendix I). Experts in evidence-based practice were asked to rate the relevance of each item on the EBPSAT on a four-point Likert scale from 1 (not relevant) to 4 (very relevant). Each item also included a comment box so that participants could provide open-ended feedback on individual items.

Demographic Questionnaires. Prior to completing the EBPSAT, participants were asked to provide personal demographic information. Items on this demographic questionnaire collected data regarding participants' gender, ethnicity, degree, most current graduation year, research courses in college, current position or title, years at current position, and professional setting. Participants who indicated they are teachers were also asked to provide the grade(s) and subjects they currently teach (see Appendix G). The demographic questions for the content validity questionnaire collected data regarding area of expertise, degree, position, professional publications, and professional presentations on the topic of evidence-based practice.

Follow-up Questionnaire. After completing the EBPSAT, participants were given the opportunity to provide feedback on the instruments itself, in the form of six follow-up questions regarding feasibility, acceptability, and usability of the measure. The questions asked about the value of the information, whether participants would change their practice based on the information, the ease and relevance of the questions, and whether they had suggestions for adding or changing the questions. The first four questions were rated on a four-point Likert-scale (1=strongly disagree to 4= strongly agree) and included comment boxes for open-ended feedback. The last two questions about adding or changing items on the survey only included open-ended response options.

Procedures

Data Collection. Following approval from the IRB, emails were posted to social media and educator message boards, as well as sent to nationwide educator listservs. The emails included a short description of the study, an embedded message that contained a link to the survey, and information about incentives for participation, including entrance into a drawing for one of two $50 Visa gift cards (see Appendix C). Participants who chose to click on the link were directed to a webpage containing the survey. This page explained the purpose of the study, the approximate time commitment for completion, and provided the researchers' contact information. The participants then chose to either complete the online survey through Survey Monkey or to exit.

According to the American Psychological Association (APA) (2010) code of ethics, when obtaining informed consent, potential participants must be informed of the following: (1) the duration, procedures, and purpose of the study; (2) participants' right to decline and withdraw from the research at any time for any reason; (3) possible consequences of withdrawal from the

study; (4) "reasonable" factors that could potentially influence willingness to participate, including risks; (5) limits to confidentiality; (6) incentives for participation; and (7) contact information for questions regarding participants' rights or about the research (APA Ethics Code, 2010). This information was provided on the first page of the Survey Monkey survey, which required participants to click "next" in order to access the questionnaire.

Data Screening. In order to ensure that the necessary underlying assumptions of the statistical analysis utilized were met, preliminary exploratory data analyses were conducted using Statistical Program for Social Sciences (SPSS). The purpose of these analyses was to detect univariate, bivariate, and multivariate practical issues including: missing or incorrectly entered data, unmet assumptions of normality, outliers, nonlinearity, multicollinearity, and singularity (Tabachnick & Fidell, 2013). The first step in this process was to identify the measured variables for analysis, which were the 20 items of the EBPSAT. The data from these items were then screened for outliers, missing data, and poorly distributed variables so that such issues could be controlled (Tabachnick & Fidell, 2013). Following screening, the data were analyzed to address each of the five research questions. The primary statistics employed were exploratory factor analysis (EFA), analysis of variance (ANOVA), and bivariate correlations.

Inaccurate and missing data. Tabachnick and Fidell (2013) state that it is necessary to compare the original data to data entered into the statistical program to ensure data accuracy. For this study, the researcher printed out the data files from Survey Monkey and visually compared them to the data imported into SPSS. The purpose of which was to screen for missing data, values that were out of range, and implausible means and standard deviations. Responses should be within the values of the Likert scale of the EBPSAT. Several strategies could be employed to deal with missing data including deleting cases or variables or assigning values to

missing cases. Selection of procedures for dealing with missing data is dependent upon the randomness and percentage of the missing data (Tabachnick & Fidell, 2013). The approach employed and its rationale are reported in the *Results* section.

Univariate and multivariate normality and outliers. Normality refers to the extent to which variables, combinations of variables, and residuals (the difference between observed values and their true value) are normally distributed. Normality was examined both statistically and graphically. Specifically, skewness (symmetry) and kurtosis (sharpness of the peak of the distribution) were examined. The data were also screened for univariate and multivariate outliers, which are extreme values that could potentially lead to Type I and Type II errors. Specifically, Mahalanobis distance, which is evaluated using a χ^2 distribution, was used to detect multivariate outliers. If significant departures from normality are detected, data transformation may be considered. Outliers were transformed or deleted. Changes to the data set are reported in the *Results* section.

Linearity, multicollinearity, and singularity. Linearity refers to the assumption of a straight-line relationship between variables. It is essential because EFA only examines linear relationships between variables. Linearity is examined using residual plots and bivariate scatterplots. According to Tabachnick and Fidell (2013), if nonlinear relationships are detected, the measured variables should be raised to a power to control for the issue. Multicollinearity and singularity refer to problems that occur when variables are too highly correlated. Several indices are used in the screening of multicollinearity and singularity including: squared multiple correlations (SMC), variance inflation factor (VIF), condition index, and variance proportions (Tabachnick & Fidell, 2013). Multicollinearity and singularity are further discussed in the *Results* section.

Data Analysis. This section will introduce the methods of analysis utilized to answer each of the five research questions following data screening. To answer the first research question regarding the underlying factor structure of the EBPSAT, exploratory factor analysis was conducted. The researcher first assessed the factorability of the items included in the instrument before examining the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. The diagonals of the anti-image correlation matrix were examined for values below 0.5. Communalities above 0.3 are expected, indicating shared variance among the items. Because the factors were assumed to be correlated, an oblique solution was expected. Exploratory factor analysis was selected because there is no hypothesized factor structure.

For the second research question, examining the internal consistency of the EBPSAT, Cronbach's Alpha reliability coefficients were calculated. According to Furr and Bacharach (2014), Cronbach's Alpha is the most widely used method of estimating reliability. As an oblique solution was expected a single reliability coefficient was calculated for the instrument. Researchers first calculated the variance of scores on the complete test, before calculating the covariance between each pair of items followed by the reliability coefficients. For the purpose of this study, reliability coefficients of 0.70-0.79 are considered "good" and 0.80 and higher are considered "excellent".

The third research question examines the content validity of the instrument. *Content validity* is defined as "the ability of the selected items to reflect the variables of the construct in the measure" (Newman, Lim, & Pineda, 2013, p. 243). The most widely reported approach for content validity is the content validity index (CVI). A panel of experts in evidence-based practice completed a content validity questionnaire, in which they rated each item's relevance on a 4-point ordinal scale (1[not relevant], 2[somewhat relevant], 3[quite relevant], 4[highly

relevant]) (Zamanzadeh et al., 2015). The CVI was first calculated for each item, by counting the number of experts who rated the item as a 3 or 4 and then dividing by the total number of experts, which will result in the proportion of experts who deem the item relevant. Next, the CVI for the entire instrument was estimated by calculating the average CVI across the items. Davis (1992) recommends correlations of 0.80 for new measures.

The fourth research question examines the construct validity of the EBPSAT. *Construct validity* refers to "the degree to which test scores can be interpreted as reflecting a particular construct" (Furr & Bacharach, 2014, p.201). In this case, the construct in question is evidence-based practice and, specifically, the three hypothesized factors of the EBPSAT: *Organization*, *Individual*, and *Practice*. According to Furr and Bacharach (2014), there is no single best way to study construct validity. Two methods of providing evidence of construct validity include the test content and the internal structure of the test, both of which have been addressed by the first two research questions.

For the purpose of this study, construct validity is addressed, in part, by the statistical analyses used to address research question 1, as previously stated. Identifying the underlying factor structure of the instrument should provide some evidence of content validity. A secondary method of examining construct validity included administering the survey to a sample of individuals who are known to implement EBP and comparing the results against a sample of teachers and administrators, in a method referred to as known-groups validity. For the purpose of this study, a convenience sample of professionals who practice applied behavior analysis, a field that is widely known to utilize EBP, specifically Board Certified Behavior Analysts (BCBAs), was used. Differences between ratings on the EBPSAT for three groups (teachers, administrators, and BCBAs) were analyzed using Analysis of Variance (ANOVA).

The fifth and final research question examines the feasibility, acceptability, and usability of the EBPSAT. Participants completed a brief questionnaire, consisting of Likert-type items related to the ease of administration, as well as the ways in which schools might use the results. Participants were also afforded the opportunity to provide comments or feedback related to each of the follow-up questions. The questions were analyzed for each of the categories of respondents (teachers, administrators, and related service providers). T-tests were utilized to determine the differences between groups. Qualitative analysis of open-ended responses was performed to identify themes among responses.

Summary

This chapter included an overview of the research design and research methodology employed in this study. The research design utilized was a multivariate correlational design primarily utilizing exploratory factor analysis, analysis of variance, and bivariate correlations. Participants were recruited using a convenience sample of individuals who responded to an electronic survey link. Data were collected using the 20-item EBPSAT, a demographic questionnaire, a follow-up questionnaire, and a content validity questionnaire administered via Survey Monkey web-based survey software.

Chapter 4

Results

The purpose of this study was to examine the psychometric properties of the newly developed Evidence-Based Practice Self-Assessment Tool. Specifically, the following questions were addressed: (1) what is the underlying factor structure of the EBPSAT? (2) What is the internal consistency reliability of the EBPSAT? (3) What is the content validity evidence of the EBPSAT? (4) What is the construct validity evidence of the EBPSAT? (5) What is the evidence of usability, acceptability, and feasibility of the EBPSAT? Results of the exploratory factor analysis were used to answer the first research question. Cronbach's alpha was used to answer the second research question. The content validity index (CVI) of the content experts survey was calculated to address the third research question. The fourth research question was partially addressed by the results of the factor analysis, as well as one-way analysis of variance (ANOVA) between a random sample of the total research sample and a sample of known evidence-based practitioners. Finally, the fifth research question regarding the usability, acceptability, and feasibility of the instrument was addressed through t-tests and qualitative analysis of open-ended responses.

Screening

Data accuracy. Prior to data analysis, the scores were examined to ensure that the data were accurately entered into SPSS. The survey responses were first downloaded from Survey Monkey as a Microsoft Excel file. The file was, then, directly imported into SPSS. Finally, the researcher visually examined both the Excel spreadsheet and the SPSS file to compare participants' responses. No discrepancies were found between the files.

Missing data. A total of 342 participants began the survey by clicking the *next* button on the informed consent page. Twenty-three participants exited the survey after responding to the demographic items. One participant exited the survey after responding to the first two survey items, twelve participants quit after completing the *Organization* section, and three participants quit before completing the *Practice* section. In total, 39 participants, or roughly 11% of the total sample that began the EBPSAT were deleted from the sample due to failing to complete the measure. Though there are no set guidelines for the allowable amount of missing data for a given sample size, any amount may affect generalizability (Tabachnick & Fidell, 2013).

At this point, 302 participants remained and their responses to the items on the EBPSAT were screened for missing values. The number of missing values for each of the 20 items ranged from 0-4. Each missing value was adjusted using mean substitution by item, which is a popular method for estimating missing data (Tabachnick & Fidell, 2013). Means were calculated using SPSS and these mean values were used to replace the missing data to enable further analysis. On the *Organization* subscale, *Don't Know* responses were treated as missing data. In total, there were twelve *Don't Know* responses across the items of the *Organization* subscale, representing a scant 4% of the remaining sample. Tabachnick and Fidell (2013) state that mean substitution is a conservative approach to handling missing data since the mean for the whole distribution does not change and the researcher does not have to guess at missing values.

Univariate and multivariate outliers. After detecting and dealing with missing values, data screening was conducted. The following describes the methods of identifying outliers and rationale for the use of such methods. When using continuous variables, cases are considered potential univariate outliers when their standardized score's absolute value exceeds 3.29 ($p < .001$, two-tailed test) (Tabachnick & Fidell, 2013). To identify univariate outliers, z-scores were

calculated for each of the sixteen continuous items of the EBPSAT. There were no z-scores exceeding 3.29, indicating a lack of potential outliers. To address univariate outliers for the four dichotomous items, the *split* between categories was examined. According to Tabachnick and Fidell (2013), cases on the "wrong" side of a very uneven split are likely univariate outliers. The authors recommend that dichotomous variables with extreme splits (90-10) be deleted. In this case the split between categories for each of the four items did not exceed 90-10.

To screen for multivariate outliers, a Mahalanobis distance statistic was calculated. The x^2 critical value using twenty degrees of freedom was, $x^2_{.999} = 45.315$. Eleven cases greater than this critical value were found: (1) 64.01, (2) 62.87, (3) 61.54, (4) 59.04, (5) 57.57, (6) 51.25, (7) 49.87, (8) 49.58, (9) 47.85, (10) 46.84, and (11) 46. 57, respectively. Item responses and demographic variables were examined to determine if these cases were representative of the target population. It was determined that these outliers were representative of the population of interest. However, as EFA is affected by non-normal data, the 11 participants who accounted for the observed outliers were deleted, leaving 291 participants.

Normality. To assess the normality of the items of EBPSAT, normality plots, histograms, and skewness and kurtosis values were examined. First, skewness and kurtosis values were transformed into z-scores to allow analysis of the distribution. Standardized skewness or kurtosis scores with absolute values that exceed 3.29 indicate a significant departure from normality (Tabachnick & Fidell, 2013). Of the twenty items on the EBPSAT, six items had standardized kurtosis values exceeding 3.29 and thirteen had standardized skewness values exceeding 3.29, indicating the presence of significant negative skewness among the items. However, the impact of skewness is diminished in large sample sizes. Additionally, while

solutions obtained in factor analysis are enhanced when variables are normally distributed, they are not dependent upon it (Tabachnick & Fidell, 2013).

The assumption of normality can also be assessed through visual inspection of normal probability plots (P-P plots). In these plots, the actual normal values are compared to the expected normal value. If the actual distribution is normal, the majority of the observed values will fall along a diagonal line with some small deviations (Tabachnick & Fidell, 2013). A visual inspection of Figure 1 provides support for normality, as each observed value falls close to the diagonal line. Figure 1 was obtained by combining the items of the EBPSAT.

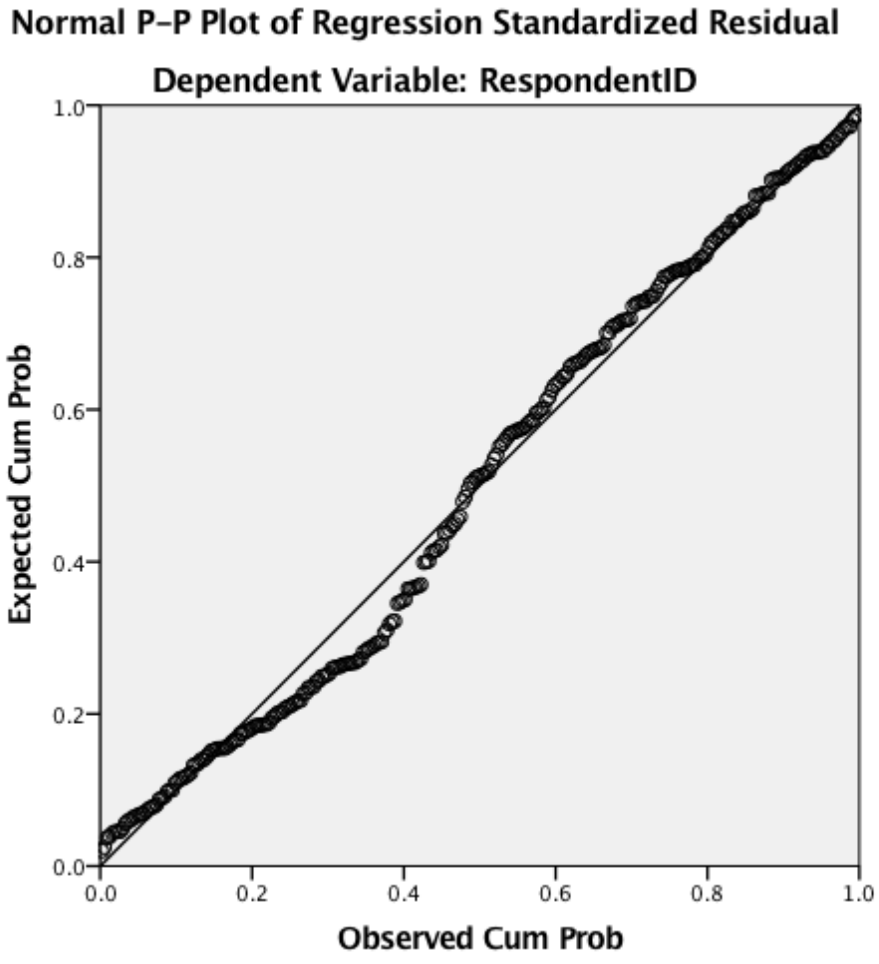

Figure 1. Normal Probability Plots of Regression—EBP Self-Assessment Tool

Another graphical representation that can be used for evaluating normality is the frequency histogram, which provides an illustration of the distribution of sample data. Histograms illustrate the extent to which the data are symmetrical (skewness) and how the data cluster around the center (kurtosis) (Tabachnick & Fidell, 2013). Visual inspection of the frequency histogram in Figure 2 provides support for normality. The data are symmetrical and generally fits the normal bell curve superimposed by SPSS.

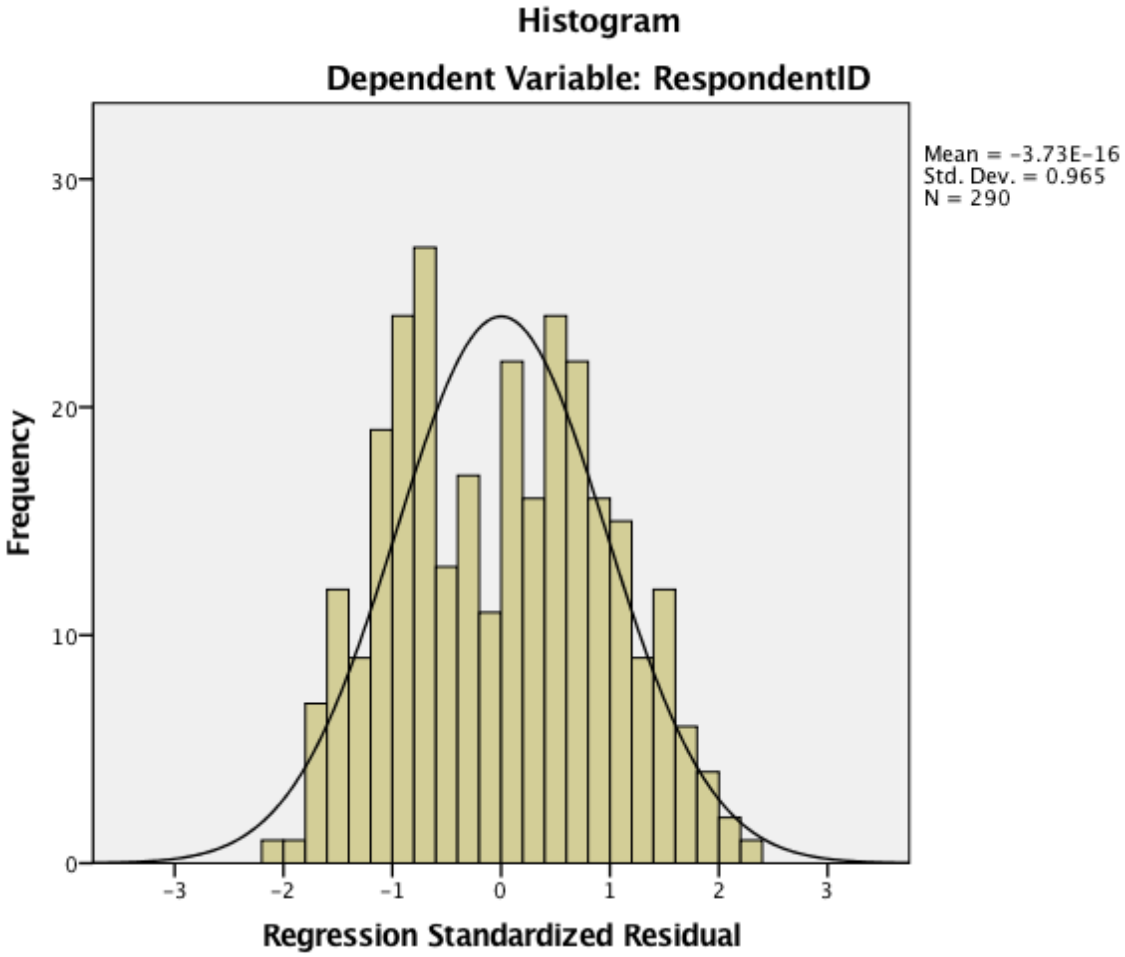

Figure 2. Histogram with Normal Curve of Responses to EBP Self-Assessment Tool

According to Tabachnick and Fidell (2013) when analyzing ungrouped data, screening the residuals is sufficient if the residuals are normally distributed. As the residual plot and

histogram look relatively normal, the screening of individual variables for normality was not necessary. There is sufficient evidence to suggest that no significant departure from normality exists and that further investigation is unnecessary.

Linearity and homoscedasticity. Linearity assumes that a relationship between variables is a straight line. As Pearson's *r* only examines linear relationships, ignoring nonlinear relationships, linearity is essential. Nonlinearity is assessed through visual inspection of bivariate scatterplots and is probable when the residuals form a curvilinear pattern. Homoscedasticity for ungrouped data assumes that the variability in a continuous variable's values is roughly equal at all values of another continuous variable's scores. Homoscedasticity is not essential, but analysis is weakened in its absence. Homoscedasticity is indicated when a scatterplot between two variables is mostly even in width with a slight bulging in the middle (Tabachnick & Fidell, 2013).

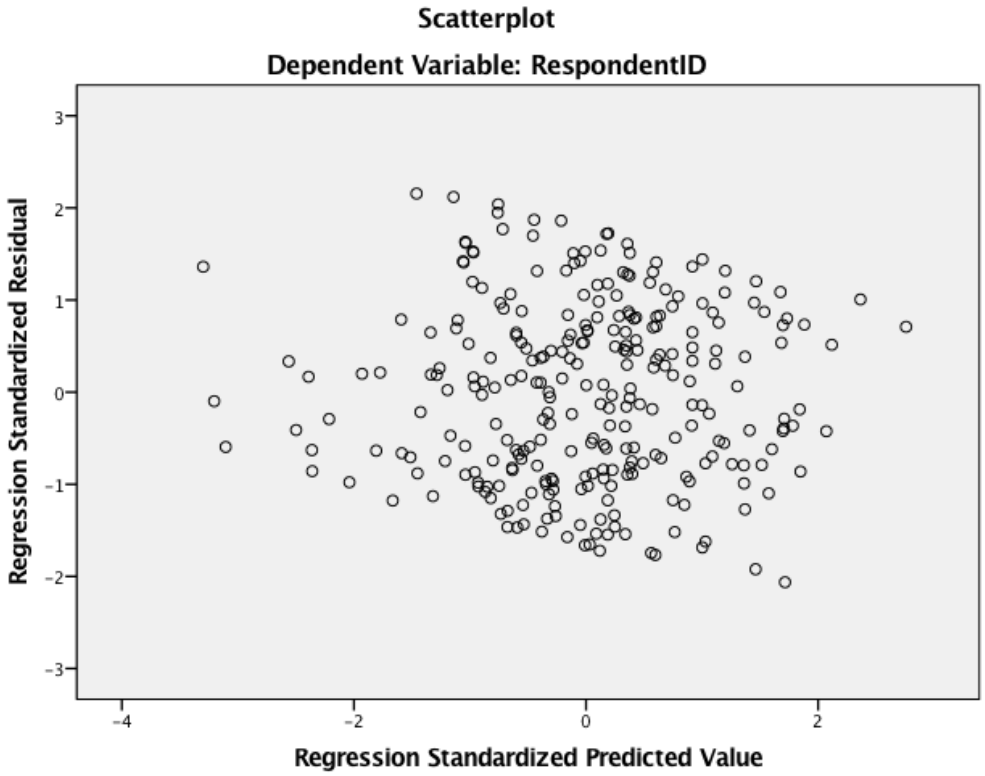

Figure 3. Scatterplot—EBP Self-Assessment Tool

Multicollinearity and singularity. Multicollinearity and singularity refer to issues that occur when variables are too highly correlated. Multicollinearity is indicated when variables are correlated above .90, suggesting that the variables measure the same construct. Singularity indicates that the variables are redundant, meaning that one variable is a combination of two or more other variables. Several indices are used in the screening of multicollinearity and singularity including: squared multiple correlations (SMC), variance inflation factor (VIF), condition index, and variance proportions (Tabachnick & Fidell, 2013).

First, SMC values were examined using tolerance values (i.e. 1-SMC) provided by SPSS. All values were found to be in the acceptable range of greater than .20.VIF values are another indicator of multicollinearity, with desirable values being less than ten (although less than 5 is preferable). All the VIF values for the 20 variables were less than 5. Condition Indices and variance proportions were the final statistics used to screen for multicollinearity. Condition Index values greater than 30 when combined with variance proportions greater than .50 indicate possible problems with collinearity. While seven dimensions had Condition Indices greater than 30, no values greater than .50 on the same dimension were found. This suggests that there is not sufficient evidence of multicollinearity according to this method.

Data Analysis

Exploratory Factor Analysis. The first research question pertains to the underlying factor structure of the EBPSAT. Exploratory factor analysis (EFA) is a statistical technique that allows researchers to discover which variables in a set are correlated with one another and, therefore, form a coherent subset of factors (Tabachnick &Fidell, 2013). If the factors are also relatively independent of one another, they are believed to be free of common underlying

processes that account for the correlation between variables. In psychology, the primary use of EFA is in the development of objective measures of mental characteristics. The four major goals of EFA are to (1) reduce a large number of variables to a smaller number of factors; (2) to concisely describe the relationship between variables; (3) to operationally define the underlying processes measured by variables; and (4) to test theories about underlying processes (Tabachnick & Fidell, 2013). EFA was used to determine the underlying factor structure of the EBPSAT.

According to Tabachnick and Fidell (2013), when using EFA, the following should occur: (1) an adequate sample size should be obtained and data screening should be conducted; (2) the factorability of *R* should be assessed; (3) factor extraction and rotation should be conducted; (4) the adequacy of extraction and number of factors should be determined; and (5) the rotated component matrix should be interpreted.

Sample size and data screening. The general rule when applying exploratory factor analysis is that a minimum of 300 cases is required. MacCallum, et al. (2001) and Thompson (2000) recommend sample sizes between 100 and 200. A sample size of at least 300 is recommended when communalities are low, there are a small number of factors, or only three or four indicators are present for each factor (MacCallum et al. 2001). The recommended ration of participants to measured variables, according to Thompson (2000) is 10:1. However, Bentler and Chou (1987) suggest a ratio of 5:1. For the purpose of this study, a sample size between 300 and 500 was attempted. The final sample size of 290 (with missing data and outliers deleted) resulted in a ratio of 15:1. As the factor loadings were moderate to high, the sample is considered to be adequate (291:20 > 5:1). Since data screening was conducted to address missing data, normality, outliers, singularity, linearity, and multicollinearity, no further exploratory data analysis is necessary.

Factorability of R. The data collected from the 290 participants was used to create a correlation matrix, which is a square, symmetrical matrix of row and columns representing each of the twenty variables. The value at the intersection of each row and column represents the correlation between two variables. At least one correlation greater than .30 should be found within the matrix in order to employ EFA (Tabachnick & Fidell, 2013). Many bivariate correlations greater than .30 were observed among the 20 items of the EBPSAT, indicating that the use of EFA was appropriate. However, the presence of a significant number of high bivariate correlations alone is insufficient proof of the existence of factors. Thus, it was necessary to examine the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity to further assess the factorability of *R* (Tabachnick & Fidell, 2013).

KMO is a measure of sampling adequacy, which is expressed as a ratio of the sum of squared correlations to the sum of squared correlations *and* the sum of squared partial correlation. KMO values greater than .60 are required for factor analysis (Tabachnick &Fidell, 2013). The KMO value for this sample was .814, indicating "meritorious" sampling adequacy (Kaiser, 1974). Bartlett's Test of Sphericity tests the null hypothesis that the correlations in a given correlation matrix are zero (Tabachnick & Fidell, 2013). If the test is significant, the null hypothesis is rejected, suggesting the presence of significant correlations in the matrix that warrant factor analysis. In this case, Bartlett's Test of Sphericity was significant and the null hypothesis was rejected, x^2 (df=190) = 2268.014, p = .00. Both the KMO and Bartlett's Test of Sphericity provide evidence for the factorability of *R* (see Figure 4).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.814
Bartlett's Test of Sphericity	Approx. Chi-Square	2268.014
	df	190

Sig.	.000

Figure 4. Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity

Adequacy of extraction and number of factors. The purpose of extraction is to estimate the parameters of the factor solution. There are two commonly used extraction methods, Maximum Likelihood (MA) and Principal Axis (PA) (Tabachnick & Fidell, 2013). For this analysis, researchers utilized principal axis factoring, which maximizes variance extracted by orthogonal factors. It estimates communalities to attempt to eliminate unique and error variance from the variables (Tabachnick & Fidell, 2013). Principal axis factoring, the most commonly used FA, is often the beginning extraction method used by researchers. Communalities, the proportion of the common variance present in a variable, were also examined. Communality values range from zero to one, with a communality of one indicating a lack of unique variance and a communality of zero indicating a lack of shared variance. Communalities in this study ranged from .032 to .858. As communalities greater than .20 are considered acceptable (Martin, 2013) three items were considered for elimination based on this criterion (Prac4= .032, Prac3= .087, and Indiv8= .182).

Eigenvalues represent the extent to which variance is accounted for by a factor and are the first indicator as to the determining the number of factors to include (Tabachnick & Fidell, 2013). Initial eigenvalues were examined using the criterion of values greater than one. Six factors emerged based on this cut-off. However, as having six factors is not parsimonious, the researcher used a more stringent criterion suggested by Tabachnick and Fidell (2013) of values greater than two, which yielded two factors. Initial eigenvalues for the two factors were, 5.108 and 2.308. The two factors accounted for 35.09% of the total variance.

Scree plots illustrate eigenvalues on the Y-axis against factors on the X-axis, which allows researchers to determine the relative importance of a factor. Generally the scree plot includes a few factors with high eigenvalues and many with low values. Of particular importance is the point of inflexion, where the slope of the line changes drastically, becoming horizontal rather than vertical (Tabachnick & Fidell, 2013). Visual inspection of the scree plot (see Figure 5) provides some evidence for a three-factor structure with a dramatic change in slope after the third factor. The total variance explained by the third factor is 6.1%, (compared to 25.61% and 9.64%) with an initial eigenvalue of 1.653, which is below the cutoff criterion of values greater than 2. This evidence along with the previously mentioned eigenvalues, suggests the existence of a two-factor structure.

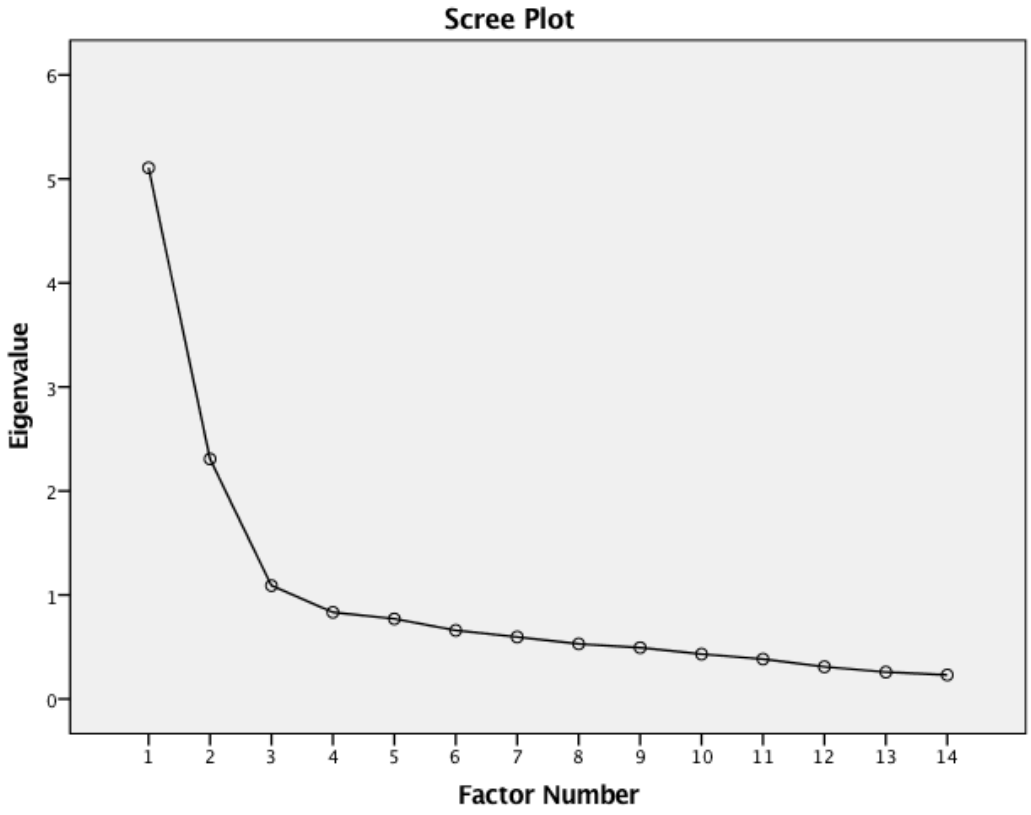

Figure 5. Scree Plot of Eigenvalues for the EBP-Self-Assessment Tool

Factor Rotation. After factors have been extracted, it is necessary to determine the extent to which the items load on these factors. Rotation is used to maximize a variable's loading on only one factor. There are two types of rotation, orthogonal and oblique. Orthogonal techniques are employed when the factors are not correlated with one another, whereas oblique rotation is utilized when there is reason to believe that the factors are correlated (Tabachnick & Fidell, 2013). Tabachnick and Fidell (2013) suggest that oblique rotation be utilized with factor correlations greater than .32. In this case the correlation between the two factors is .429. Thus, oblique rotation was used. SPSS provides two methods of oblique rotation, direct oblimin and promax. Promax rotation is intended for larger sample sizes and is quickly completed, whereas direct oblimin allows for a wider range of intercorrelations, and may take longer (Tabachnick & Fidell, 2013). Direct oblimin rotation was utilized due to the relatively small sample size of the current study. The delta value, which is a parameter that "controls" the extent of obliqueness among factors, was set to the default of 0. Using direct oblimin, the factor structure was found to be moderately oblique, as evidenced by the moderate correlation between factorsas (r=.429) (see Table 4).

Table 4

Correlations Between the Two Factors of the EBP Self-Assessment Tool

Factor Correlation Matrix

Factor	1	2
1	1.000	.429
2	.429	1.000

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin with Kaiser Normalization

Interpretation. Interpretation of factors is the final step in factor analysis. This allows researchers to understand the dimension of unifying group loadings and the amount of

covariance explained by the factors (Tabachnick & Fidell, 2013). The total variance explained by the two factors was 35.09%. Table 2 provides more details of the variance explained by the two factors of the EBP Self-Assessment Tool.

Table 5

Extraction Sums of Squared Loadings for the Two Factors: Explained variance for each factor.

Factor	Total	% of Variance	Cumulative %
1	5.10	25.52	25.52
2	1.91	9.56	35.09

Note. N=291. Extraction Method: Principal Axis Factoring.

When interpreting item loadings on each factor, the minimum criterion for meaningful loadings is .32. However, as this correlation is considered to be "poor", the criterion suggested by Tabachnick and Fidell (2013) of .45 was used. This criterion is considered "fair". All five of the items on the *Organization* subscale loaded at .45 or higher and were retained. Two of the eleven items that comprise the *Individual* subscale failed to meet the criterion and were suppressed. All four items comprising the *Practice* subscale loaded below .45 and were also suppressed. Of the sixteen remaining items, none were cross-loaded, and loaded on only one factor.

The two factors represent two of the three original subscales: *Individual* (Factor 1) and *Organization* (Factor 2). Ten of the eleven items from the *Individual* subscale loaded on factor one, and the five items from the *Organization* subscale loaded on factor two; with the addition of question 8 from the *Individual* subscale. However, upon examination, it is evident that question 8 (*I have sufficient resources and support to implement EBP)* is related to EBP at the organizational level, rather than the individual level. A more exhaustive interpretation of these

results and their implications will be presented in the *Discussion* chapter. More information regarding the factor loadings and the corresponding items are reported in Table 3 and Table 4.

Table 6

Status of Item Retention by Subscale

Item	Question	Subscale	Retained
Q1	EBP is central to the mission and philosophy of my organization.	Organization	Yes
Q2	Monitoring and sharing outcomes are essential in my organization's culture.	Organization	Yes
Q3	Fiscal resources are used to support the implementation of EBP.	Organization	Yes
Q4	Administrators regularly supply resources (e.g. articles, reports, professional development) to support the implementation of EBP.	Organization	Yes
Q5	EBP is implemented in my organization.	Organization	Yes
Q6	I value EBP as a decision-making framework.	Individual	Yes
Q7	I have sufficient reason or motivation to implement EBP.	Individual	Yes
Q8	I have sufficient resources and support to implement EBP.	Individual	Yes
Q9	I have sufficient time to implement EBP.	Individual	No
Q10	When making decisions, I value the characteristics, values, and preferences of my client(s).	Individual	Yes
Q11	When making decisions, I value clinical/field experience.	Individual	Yes
Q12	When making decisions, I value research evidence.	Individual	Yes
Q13	I regularly use professional resources other than research studies (e.g. websites, practice guides) to inform my practice.	Individual	No
Q14	I regularly use client data to inform my practice.	Individual	Yes
Q15	I regularly use research findings to inform my practice.	Individual	Yes
Q16	I have an ethical and professional responsibility to implement EBP.	Individual	Yes
Q17	In the past eight weeks I have read a research study.	Practice	No
Q18	In the past eight weeks I have shared findings of a research study with a colleague, administrator, or client/family member.	Practice	No

Item	Question	Subscale	Retained
Q19	In the past eight weeks I have shared progress data with a colleague, administrator, or client/family member.	Practice	No
Q20	In the past eight weeks I have collected data to examine the impact of practice.	Practice	No

Table 7

Pattern Matrix of 14 Variables

Item	Factors	
	1	2
Q7	.773	
Q16	.732	
Q6	.728	
Q13	.637	
Q10	.597	
Q15	.592	
Q14	.556	
Q11	.490	
Q4		.773
Q1		.769
Q3		.713
Q5		.688
Q2		.682
Q8		.534

Note. Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 4 iterations.

Internal Consistency Reliability. The second research question pertains to the internal consistency reliability of the EBPSAT. Cronbach's alpha statistics were calculated for the entire EBPSAT (20 original items), as well as for the three original subscales, to determine the degree

of internal consistency reliability. The coefficient alpha for the total test (20 original items) was .864 and the coefficient alpha's for the three original subscales were: *Organization* (.860), *Individual* (.838), and *Practice* (.533). Reliability statistics were then computed for the entire scale with the 14 remaining items, based on the results of the EFA, as well as for the two remaining subscales *Individual* and *Organization)*. Item 8 from the *Individual* scale was included on the *Organization* scale for the purpose of this analysis. The coefficient alpha for the total test (14 remaining items) was .857, with the coefficient alphas for the two factors being: *Individual* (.849) and *Organization (*.849). George and Mallery (2003) offer the following criteria: coefficient alphas greater than .90 indicate "excellent" inter-item correlations, while values of .80-.90 are considered "good". Using these criteria, the coefficient alphas suggest that the subscales and items are consistently measuring the underlying constructs. Additionally, a moderate correlation was found between the *Individual* and *Organization subscales (r*= .447). However, weak correlations were found between the *Individual* and *Practice* subscales (*r*= .423), and between the *Organization* and *Practice* subscales (*r*= .147). Finally, with the elimination of the six items suggested in the EFA, a moderate correlation between the remaining factors was found (*r*=. *429).* The implications of these findings will be further discussed in the Discussion chapter.

Content Validity Index. Research question three examines the content validity evidence of the EBPSAT. Twelve experts completed the content validity questionnaire, rating items on the EBPSAT on a four-point Likert scale (1=not relevant, 2= somewhat relevant, 3=relevant, 4=very relevant). To estimate the content validity index for each item (I-CVI), the number of experts who rated the item as either a three or four were counted and divided by the number of experts. The CVI for the entire survey (S-CVI) was calculated by calculating the average CVI for all

items. According to Polit and Beck (2006), when a panel consists of five or fewer experts, all experts should agree on an item's relevance for their rating to be considered a "reasonable representation" of possible ratings (p.491). This method, referred to as universal agreement, is more stringent than using the average item CVI, however, with greater than 6 experts the likelihood of chance disagreement increases exponentially, making an average CVI more appropriate (Polit & Beck, 2006).

The CVI for the twenty items ranged from .67 to 1.00. Two items have a CVI of .67; twelve items have a CVI of .92; and the remaining six items have a CVI of 1.00. The average of these is .92, which is the CVI for the scale and is well above the .80 criteria. Though two item scores fell below the .80 criteria, it should be noted that both items were on the *Practice* subscale, which was recommended for deletion during the factor analysis. Table 5 illustrates the expert ratings and CVIs for the EBPSAT.

Table 8

Item Content Validity Indices

	Expert												
Item	1	2	3	4	5	6	7	8	9	10	11	12	CVI
1	4	4	4	3	3	4	4	**2**	4	3	4	4	.92
2	4	4	4	4	4	4	4	4	3	4	3	4	1.00
3	3	4	4	4	4	3	4	**2**	3	4	3	4	.92
4	4	4	4	4	4	4	3	4	3	4	**2**	3	.92
5	4	3	4	3	4	4	4	4	3	4	3	4	.92
6	4	4	4	4	4	4	2	4	3	3	4	4	.92
7	4	4	4	4	**2**	4	4	4	3	4	4	4	.92

8	4	3	3	4	4	3	3	**2**	3	4	4	4	.92
9	3	4	4	4	4	4	3	3	3	4	4	4	.92
10	4	4	4	4	**2**	3	4	4	4	4	4	4	.92
11	4	4	4	4	3	3	3	4	4	3	4	4	1.00
12	4	4	4	4	4	3	4	4	4	4	4	4	1.00
13	3	4	4	4	4	3	4	**2**	3	3	3	4	.92
14	4	4	4	4	4	4	4	4	4	4	4	4	1.00
15	4	3	4	4	4	4	4	4	4	4	4	4	1.00
16	4	4	4	4	4	4	4	4	4	3	4	4	1.00
17	3	**2**	4	3	**2**	4	4	4	**2**	**2**	4	3	.67
18	3	4	4	**2**	**2**	4	4	4	**2**	**2**	4	3	.67
19	4	**1**	4	4	4	4	4	4	**2**	**2**	4	3	.92
20	4	**1**	4	4	4	4	3	4	4	4	4	4	.92
												Scale CVI=	.92

Known Groups Validity. Known-groups comparisons can provide evidences of construct validity of a survey instrument when the members of groups with known characteristics obtain scores that are hypothesized a priori for those groups. For the purpose of the current study, three job category groups: BCBAs, teachers, and administrators were compared to other respondents in the data set. The hypothesis included that BCBAs would obtain higher overall scores on the EBPSAT than both teachers and administrators. No significant difference between teachers and administrators was hypothesized.

A one-way analysis of variance (ANOVA) was conducted to evaluate the effects of profession (BCBA, teacher, administrator) on scores on the EBPSAT. The means and standard

deviations for the scores on the EBPSAT as a function of the three professions are presented in Table 9.

Table 9

Means and Standard Deviations of EBP-Self Assessment Tool Scores by Profession

	Score on EBP Self-Assessment Tool							
	Organization Total		Individual Total		Practice Total		Total Score	
Profession	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
BCBA	17.55	2.45	41.20	1.99	3.65	.67	63.80	3.38
Teacher	13.20	3.30	37.15	4.10	3.00	1.03	55.30	7.05
Administrator	14.60	3.11	36.40	4.23	3.45	.83	56.25	6.41

The test for normality, examining standardized skewness and the Shapiro-Wilks test indicated that the data were statistically normal. The test for homogeneity of variance was not significant, *Levene* F (2, 57)= 2.809, p= .069, indicating that the assumption underlying the application of the ANOVA was met. An alpha level of .05 was used for the initial analyses. The results for the one-way ANOVA indicate a significant main effect for profession and Total Score F (2, 59)= 12.73, p< .001; *Organization* Total F (2, 59)= 11.09, p< .001; and *Individual* Total F (2, 59)= 10.32, p< .001. The effect of profession on *Practice* Total was not found to be significant, F (2, 59)= 3.05, p= .06.

Table 10

One-way Analysis of Variance for EBP Self-Assessment Tool Scores by Profession

Source	*SS*	*df*	*MS*	*F*	*p*
Organization Total	197.23	2	98.62	11.09	.000
Individual Total	266.70	2	133.35	10.32	.000
Practice Total	4.43	2	2.21	3.05	.055
Total Score	867.70	2	433.85	12.73	.000

Note. SS= sum of squares, df= degrees of freedom, MS= mean square

Tests of the a priori hypotheses were conducted using Bonferroni adjusted alpha levels of .017 per test (.05/3). Results indicated that Total Score was significantly higher for the BCBA group (M= 63.80, SD=3.8) than for teachers (M=55.30, SD=7.05) and administrators (M=56.25, SD= 6.41). However, no significant differences were found between teachers and administrators. Taken together, these results indicate that the EBPSAT can differentiate between groups of practitioners. Specifically, our results confirm that BCBAs report more use of evidence-based practice at both the organizational and individual level than do teachers and administrators.

Feasibility, Acceptability, and Usability. To address the feasibility, acceptability, and usability of the EBPSAT, participants responded to six follow-up questions, four of which included four-point Likert-type responses (1=strongly disagree to 4=strongly agree) and two of which were open-ended. Each of the four Likert-type questions included a comment box, so that participants could provide open-ended feedback as well.

Feasibility. Feasibility refers to the ease of completing the EBPSAT. To address this question, participants rated the statement, "The questions on this survey were relevant to my practice and easy to answer" from *strongly disagree* to *strongly agree.* Of the 291 participants,

50% (n=145) responded *somewhat agree,* 43% responded *strongly agree* (n=124), 6% (n=20) responded *somewhat disagree,* and 0.7% (n=2) responded *strongly disagree.* The mean rating was 3.35 with a standard deviation of 0.64.

In general, the open-ended comments reiterated that the questions were, in fact, easy to answer. One participant noted that the question is "double barreled" and should be divided into two separate questions ("The questions on this survey were relevant to my practice" and "The questions on this survey were easy to answer"). Two participants suggested changing the terms "practice" and "clients" to "teaching" and "students". Finally, one participant indicated the need for clarification boxes for three of the items ("When making decisions, I value the characteristics, values and preferences of my client(s)"; "When making decisions I value clinical/field expertise"; and "I regularly use client data to inform my practice").

Acceptability and Usability. Acceptability and usability refer to whether participants find value in the survey instrument and the usefulness of its results. To answer this question, participants rated the remaining three follow-up questions: (1) I think the information collected by this survey is valuable; (2) I think the data from this survey will be useful in my workplace; and (3) I will change my practice after taking this survey. Regarding the value of the information, 55% (n=159) of participants responded *somewhat agree,* 39% (n=113) responded *strongly agree,* 5% (n=16) responded *somewhat disagree,* and 1% (n=3) responded *strongly disagree.* The mean rating was 3.31 with a standard deviation of 0.62.

The open-ended responses regarding the value of the information fell primarily into three categories: *valuable, not sure,* and *not valuable.* Four participants indicated that they would be interested in the results of the current study and three others indicated that they were not sure how the information could be used. Of those that indicated that the information was not valuable,

two indicated problems with survey research in general, one stated that the concept of EBP was not well-defined in the current study, and one responded, "Leave education to the teachers and administrators that work hands on with kids every day. Get the companies out of the classroom. Get those that don't work with kids away from deciding what works for children. They don't".

Regarding the usefulness of the data, 65% (n=190) of participants responded *somewhat agree,* 20% (n=58) rated the item *strongly agree,* 12% (n=36) responded *somewhat disagree,* and 2% (n=7) responded *strongly disagree.* The mean rating was 3.03 with a standard deviation of 0.65. The open-ended responses for this question, again, fell into three categories: *useful, not sure,* and *not useful.* Among those that found the survey information useful, many stated that they believed EBP was important, but that it is not used enough in schools. Additionally, several participants indicated that they would like their staff to become more familiar with EBP. Those in the *not sure* category primarily indicated that they did not have enough information to decide if the information would be useful. In general those that indicated that the information would not be useful primarily cited barriers to implementing EBP in their settings. Several participants indicated that their administration does not support the implementation of EBP. One participant indicated that the definition of EBP that was used for the current study limits the usefulness of the data.

Finally, regarding whether participants will change their practice, 48% (n=139) responded *strongly disagree*, 37% (n=107) responded *somewhat disagree*, 16% (n=45) responded *strongly agree*, and 2% (n=6) responded *somewhat agree.* The mean rating was 2.86 with a standard deviation of 1.04. Those that provided open-ended responses fell into four categories: those who already practice EBP, those who will change their practice, those who will change their thinking, and those who will not change their practice. Of the participants who

indicated that they would change their practice, four mentioned barriers to EBP implementation including: difficult administration, district-wide policies, and student characteristics. One participant stated,

> The challenge for a school principal is getting buy in from teachers or parents when the research conflicts with a practice they are familiar with and comfortable using (like homework). In a union environment, teachers have a lot of latitude to accept or reject evidence-based practices, even when collaboration time is devoted to sharing research-based practices.

Three participants stated that they would not change their practice, but that they would have increased awareness of EBP, while one participant responded, "I feel guilty for not reading enough". The final group of individuals who indicated that they would not change their practice was the largest of the four groups. In general, the responses indicated that completing a survey was insufficient for changing their behavior.

Summary

To summarize the results of this study, exploratory factor analysis was utilized to answer the first two research questions regarding the underlying factor structure and internal consistency reliability of the EBPSAT. Results of the EFA yielded a two-factor structure, with 14 of the 20 original items loading onto one of the two factors. Additionally, high Cronbach's alpha values were found, suggesting that the items are consistently measuring underlying constructs. To answer the third research question, the content validity index was calculated for each item, as well as for the entire survey. Item CVI values were consistently high (.67 to 1.00), with the exception of two items on the *Practice* subscale. The average CVI for the survey as a whole was also high (.92).

The fourth research question examined the construct validity of the instrument using a *known groups* approach. Results of a one-way ANOVA yielded significant main effects for profession (BCBAs, teachers, and administrators) and Total Score, *Organization* total, and *Individual* total. No significant effects were found for *Practice* total. Additionally, significant differences were found between BCBAs and teachers and administrators. Group differences between teachers and administrators were not significant. The final research question relates to the feasibility, usability, and acceptability of the EBPSAT. Examination of means and standard deviations of the follow-up questions indicated that participants generally found the instrument to be easy to complete and that the information held some value.

Chapter 5

Discussion

The previous chapter provided the results of EFA, internal consistency reliability, content validity index, and one-way ANOVA to provide evidence of the psychometric properties of the EBPSAT. Results of the EFA indicated a two-factor structure, with fourteen of the original twenty items retained. This model resulted in good internal consistency reliability coefficients. The CVI was calculated for each item and for the entire scale, resulting in moderate evidence of content validity. Results of the one-way ANOVA indicated that the EBPSAT is sufficiently able to discriminate between groups, suggesting evidence of construct validity, which is also supported by the results of the EFA. Additionally, participant ratings provided evidence of feasibility, usability, and acceptability of the instrument. This chapter will further discuss these results in the following order: (1) overview of the study; (2) results of each of the five research questions; (3) limitations and considerations; (4) implications for schools; (5) suggestions for future research; and (6) summary and conclusions.

Overview of the Study

Recent additions to federal and state legislation emphasize the necessity of implementing evidence-based practice in school settings. Most recently, the *Every Student Succeeds Act* (2015) outlined a definition of *evidence-based* in terms of research evidence, resulting in three levels of "acceptable evidence": Tier I or "strong evidence", Tier II or "moderate evidence", and Tier III or "promising evidence". However, the definition provided by ESSA is too narrow, conceptualizing EBP as a noun, referring to specific interventions or practices. For the purpose of the current study, EBP is defined as a decision making process, which is informed by three sources: (1) best available research evidence, (2) clinical expertise, and (3) client and family

characteristics, values, and preferences. In this definition "practice" refers to all professional activities of an individual and not a single intervention or program.

While there are legislative mandates to implement EBP in schools, there is little evidence to suggest that it is being done. To further compound the issue, there is a paucity of instruments intended to assess barriers, attitudes, and use of EBP in schools. The literature cites a multitude of reasons for teachers to avoid research and research-based approaches, including: an overall negative perception of research (Cook, Cook, & Landrum, 2013; Gore & Gitlin, 2004), the belief that research is not applicable to classroom-specific contexts (Cain, 2016; Gore & Gitlin, 2004), valuing information and insight from colleagues over the literature (Landrum, et al., 2002), and an inability to access the research (Levin, 2004; Nealet al., 2015; Rey & Gaussel, 2016; Vanderlinde & van Braak, 2010).

The current study sought to introduce a new instrument to assess attitudes toward and use of EBP in schools and investigated the psychometric properties of the instrument, as well its feasibility, acceptability, and usability. Twelve content experts rated the relevance of the items on the EBPSAT and provided open-ended feedback about each item to provide evidence of content validity. Content validity index (CVI) scores were calculated for each item and for the entire scale. Of the 342 respondents, 291 were retained for analysis, following deletion of missing data and multivariate outliers. Exploratory factor analysis was conducted to examine the underlying factor structure of the EBPSAT and internal consistency reliability correlation coefficients were calculated for the entire scale and each of the subscales. Known-groups validity was investigated by comparing EBPSAT scores for three groups: teachers, administrators, and BCBAs to determine if the instrument could discriminate between groups. Finally, feasibility, acceptability, and usability of the instrument were assessed by examining means and standard

deviations of the four Likert-type follow-up questions and assessing respondents' open-ended responses. The results of each of these analyses are discussed further in the following sections.

Preliminary Psychometric Properties of the EBPSAT

The primary objective of the current study was to investigate the psychometric properties of the EBPSAT, including: internal structure, reliability, and validity evidence. The preliminary analysis provided some evidence of internal consistency reliability, as well as evidence of content validity and construct validity. However, EFA resulted in a two-factor structure, rather than the three-factor structure, that was hypothesized, necessitating the deletion of an entire subscale. The implications of these analyses are discussed further in the following sections.

Factor structure. The first research question states, "What is the underlying factor structure of the EBPSAT?" To address this question, exploratory factor analysis was conducted. The EFA resulted in two primary changes from the originally proposed dimensionality of the EBPAS. The most significant of which, was the deletion of the *Practice* subscale, as the four items had the lowest initial factor loadings or failed to load on either of the two factors. The items on the *Practice* subscale, unlike the other two subscales, were rated dichotomously (yes/no). While the research indicates that binary items can be factor analyzed using conventional methods, the results of such analysis may be difficult to interpret (Hahs-Vaughn, 2016). For example, the results may reflect the endorsement rate of the variables rather than the construct in question, which, in this case would be EBP. In future revisions of the scale, researchers should consider altering the format of the *Practice* subscale or utilizing a specialized factor analytic program for data analysis. Two additional items with low factor loadings were deleted from the *Individual* subscale, while another item ("I have sufficient resources and support to implement EBP") loaded on the *Organization* factor, rather than the *Individual* factor,

where it was originally placed. Based on the results of the EFA, the updated survey consisted of 14 items: six on the *Organization* subscale and eight on the *Individual* subscale. More on the implications of these findings will be discussed in *suggestions for future research*.

Internal consistency reliability. The second research question states, "What is the internal consistency reliability of the EBPSAT?" Internal consistency reflects "the coherence (or redundancy) of the components of a scale" (McCrae, Kurtz, Yamagata, & Terracciano, 2011, p. 29). It was measured using the coefficient alpha or Cronbach's alpha, which is the most widely reported reliability coefficient. The coefficient alpha for the total test, as well as for each of the subscales fell in the "good" range, with each of the values being greater than 0.80. This suggests that the items of the EBPSAT are, indeed, measuring the same construct. However, while coefficient alpha can provide evidence of survey reliability, it is not sufficient on its own.

Internal consistency reliability can be impacted by content heterogeneity, which refers to whether scale items cover many aspects of a construct or a limited number (McCrae et al., 2011). A scale with a few items covering many factors will have lower internal consistency than one in which several items assess fewer factors. Additionally, item relevance (or irrelevance) can affect internal consistency reliability, which makes consistency analyses useful in eliminating irrelevant or "bad" items during scale development. However, it should be noted that internal consistency reliability is a characteristic of scores from the sample population, rather than a characteristic of the test itself since tests themselves are neither reliable nor unreliable (Green, Chen, Helms, & Henze, 2011). Thus, further reliability assessment should be conducted utilizing samples of the target population.

Content validity evidence. The third research question states, "What is the content-validity evidence of the EBPSAT?" According to Furr and Bacharach (2008) a test with

adequate content validity evidence should contain no items that are irrelevant to the construct while including, to the extent possible, the full range of content pertinent to the construct. An extensive literature review was conducted to determine the content that should be included in the EBPSAT, the objective being to include items to address attitudes and barriers related to EBP at the organizational level and the individual level, while also incorporating items related to actual use of EBP in practice. However, literature reviews are not sufficient measures of content validity themselves.

The best way to ensure sufficient content validity evidence is to enlist experts in the field and have them evaluate the test items (Furr & Bacharach, 2008). Thus, the content validity index was calculated utilizing expert ratings of the relevance of each of the items on the EBPSAT. The results indicated adequate CVI values for the majority of the individual items, as well as for the entire scale. There were two items with CVI values below. 0.80 however, both items were deleted from the final survey structure during the EFA.

While the CVI values provide evidence of content validity, this only addresses the relevance of the survey items, without addressing the second criterion mentioned by Furr and Bacharach (2008): the range of pertinent content covered by the instrument. Future studies should attempt to address this factor. Specifically, in the current study, feedback from the experts and the research participants indicated that the EBPSAT should include more items related to barriers to EBP implementation.

Construct validity evidence. The fourth research question states, "What is the construct-validity evidence of the EBPSAT?" Known-groups methods were employed to provide evidence of construct validity, which refers to "the degree to which an instrument measures the trait or theoretical construct that it is intended to measure" (Bolarinwa, 2015, p.

197). In this case, if a measure is purported to assess a particular trait, it should be able to distinguish between subsets of the population known to possess that trait and those with whom the trait is not yet established. For the current study, scores on the EBPSAT were compared for three professions: teachers, administrators, and BCBAs; the hypothesis being that scores for BCBAs would be significantly higher than those of teachers and administrators.

Results of the one-way ANOVA found significant differences between BCBAs and teachers and BCBAS and administrators. However, no significant differences were found between teachers and administrators. This indicates that both teachers and administrators report similar attitudes toward and use of EBP. Additionally, BCBAs' scores for both the *Organization* subscale and the *Individual* subscale were found to be significantly greater than those of teachers and administrators. However, no significant differences were found for the *Practice* subscale. These results suggest that BCBAs tend to be employed in organizations that are more supportive of EBP than those of teachers and administrators, in addition to being more receptive to EBP at the individual level. The EBPSAT's failure to detect differences between the three groups on the *Practice* subscale may be attributed to the fact that the subscale only consisted of four items, which were rated dichotomously. Inclusion of a larger number of items, may improve the instrument's ability to detect differences in practice of EBP. Additionally, the wording of the practice items, may have affected participants' responses. Some participants suggested that words such as "clients" and "practice" be replaced with "students" and "teaching".

Feasibility, Acceptability, and Usability

The fifth research question states, "What is the evidence of feasibility, acceptability, and usability of the EBPSAT?" When developing assessment instruments for use with particular populations, it is important to consider input from individuals from the target population. An

extensive literature review was conducted to identify factors that would increase the feasibility, acceptability, and usability of the EBPSAT for teachers, administrators, and school-based related services providers. The results of the literature review indicated that the instrument should be short, no more than a single page; simple so that it is easily understood; and practically relevant to participants' day-to-day experiences.

Analysis of the follow-up questions provided some evidence of feasibility, acceptability, and usability. The majority of respondents indicated that they found the items on the EBPSAT to be relevant and easy to answer (50% rated *somewhat agree* and 43% rated *strongly agree*). Respondents also indicated that they found some value in the information collected by the EBPSAT (55% rated *somewhat agree*) and that they thought the information would be useful (65% rated *somewhat agree).* While respondents generally rated the follow-up questions positively, the majority of indicated that taking the survey alone was insufficient to change their practice. The literature indicated that teachers and administrators are more likely to complete surveys if they are practically relevant to their day-to-day lives (Gore & Gitlin, 2004). This was evidenced by open-ended feedback from several respondents, which suggested that the items on the EBPSAT be worded specifically for teachers, administrators, and school-based professionals by replacing the terms *practice* and *clients* with *teaching* and *students.* Researchers should consider this feedback in future revisions of the EBPSAT.

Limitations and Considerations

There are several possible limitations to this study that may have affected the results and their generalizability. First, the sample size for the EFA fell slightly short of the minimum 300 that was attempted. Additionally, over 50% of the individuals in the sample had advanced degrees (Masters degree, professional degree, or Ph.D.). This may have skewed the data as it is

not representative of the general population of individuals working in schools. Also, there were significantly more administrators and school-based related service providers in the sample than teachers. When conducting future studies, a larger sample of teachers and individuals without advanced degrees is recommended.

Another potential limitation is that the content validity analysis was conducted simultaneously with the EFA and reliability analysis, rather than prior to the other analysis. Generally, in instrument development, expert analysis of the survey items is conducted prior to reliability and validity analysis so that revisions can be made based on the experts' feedback. In the current study, expert feedback was supported by the EFA, which eliminated the items found to be least relevant by the experts. Additionally, the content validity questionnaire only addressed items relevance. It did not assess item clarity or how well the construct of EBP was addressed by the survey items. Future research should examine different aspects of the survey questions. Finally, generally when calculating the CVI, researchers include both content experts and lay experts, which are laypersons who provide feedback on the survey items as well. While this is similar to assessing feasibility, acceptability, usability, it provides quantitative information about item relevance from another perspective.

The final potential limitation is due to the reliance of self-reports. As the EBPSAT is a self-report measure, participants may have chosen not to answer questions, misinterpreted items, or otherwise provided inaccurate responses. For example, social desirability, the phenomenon in which respondents provide responses they believe to be culturally appropriate rather than reporting their true beliefs, may have led to disingenuous responses (Crowne & Marlowe, 1960). As one participant stated in their feedback, "Who is going to say that they're not doing this?"

This is a difficult limitation to address, however, ensuring participants that their responses are anonymous can ameliorate it.

Implications for Teachers, Administrators, and Related Service Providers

The results of the current study provide some evidence of reliability and validity of the EBPSAT. Specifically, if properly validated, it would provide a useful tool for school districts, as there is an increasing call for accountability in education. Though the term *evidence-based practice* is fairly new to the literature, having been introduced in the mid 90's to medicine, it has steadily gained traction. Now, EBP is a buzzword, not just in the medical field, but also in social work, mental health, and education. Having recently debuted in federal education legislation, EBP is no longer just a good idea. It's the law. However, the law provides little guidance for school districts regarding accountability.

According to Maheady and colleagues (2016), "the identification, implementation, and sustainability of EBP are complicated processes that pose "wicked problems" for educational reformers (p. 109). Further, teachers and school personnel are often resistant to such processes. As a group, teachers tend to resist change, especially when it relates to their teaching practices (Jones, 2009; Landrum et al., 2007; Vanderlinde & van Braak, 2010). Many teachers do not trust research, claiming that it is not relevant to their day-to-day experiences. Thus, when faced with a problem or novel situation, teachers are more likely to seek the advice of a colleague or rely on what they learned in school, rather than turning to the research (Landrum et al., 2007). Moreover, teachers' lack of training, as well as the cost of journal subscriptions, may prevent them from being able to access the research (Levin, 2004; Neal et al., 2015; Rey & Gaussel, 2016; Vanderlinde & van Braak, 2010).

Sahin and Yildirim (2016) identify three factors that affect teacher change: school context, individual factors, and staff development factors. Where school context includes socioeconomic status, teacher workload, and administrative support; individual factors can encompass motivation, self-efficacy, and self-reflection; and staff development factors are related to the specific content and activities of professional development (PD). Gulamhussein (2013), however, argues that the majority of PD is ineffective in effecting teacher change. She offers five principles of effective PD: (1) the duration of PD should be significant and ongoing; (2) teachers must have support during the implementation stage; (4) teachers' exposure to new concepts should be active rather than passive; and (5) concepts and strategies should be modeled. Additionally, district PD decisions should be informed by data regarding gaps in knowledge and/or practice to address the needs of the staff. The EBPSAT has the potential to provide such data.

Suggestions for Future Research

As discussed throughout this chapter, the EBPSAT shows promise as a useful tool for informing school districts' professional development decisions. However, further research using the EBPSAT needs to be conducted. Specifically, research should be carried out that revises the instrument based on the results EFA, assesses the comprehensiveness of the instrument, and provides further validity evidence of the revised instrument.

The EFA from the current study resulted in a two-factor structure with 14 of the 20 original items being retained. Additionally, the EFA resulted in the deletion of the *Practice* subscale. The content-validity analysis also resulted in low CVI values for two of the four *Practice* items, indicating that the experts did not find the items to be particularly relevant. This could be related by the way in which the items were worded. The items asked respondents to

indicate whether or not they had completed EBP activities in the past eight weeks, including: reading a research study, sharing findings of a research study, sharing progress data, and collecting data to examine the impact of practice. These items may not have been worded in such a way that teachers and school-based professionals could relate to them. For example, the term *progress data* could be substituted with *progress monitoring data*, which is a term commonly used in schools. Additionally, several participants indicated that the terms *client* and *practice* be replaced with *student* and *teaching* to better relate to those who work in schools. Thus, future revisions to the EBPSAT should consider altering item wording.

For the purpose of the current study, the CVI was calculated based on expert ratings of item relevance. However, the questionnaire did not address item clarity or the comprehensiveness of the items. Content-validity addresses both the relevance of survey items and the comprehensiveness of the content (Furr & Bacharach, 2008). Feedback from participants of the current study indicated that the instrument should include more items to assess barriers to EBP implementation. As previously discussed, the items on the *Practice* subscale did not load on either of the two factors in the EFA, nor did they contribute to a third factor, as the researchers had anticipated. However, it is important not just to assess attitudes toward EBP, but actual use of EBP. Future studies should revisit the *Practice* subscale, perhaps by integrating practices more relevant to teachers' experiences.

Finally, future studies should further assess the validity evidence of the EBPSAT. One method of doing so would be to utilize an entirely new sample. EFA requires a minimum sample size of 300, slightly larger than the sample that was obtained in the current study. The current study also utilized convenience sampling for data collection, which resulted in a significant proportion of participants with advanced degrees. Future studies should consider more purposive

sampling methods to obtain samples that are more representative of the target population, thus increasing external validity. Specifically, a larger number of teachers should be recruited. Criterion validity evidence, which is the degree to with a test's scores allow for differentiation of groups (Furr & Bacharach, 2008), would also be an important area of future research. One method of addressing this would be to compare participants' scores on the EBPSAT with scores on a parallel measure of EBP. However, at the time of the current study, a parallel measure could not be identified.

Summary and Conclusions

Education has entered an era of high stakes accountability, in which, federal legislation mandates the use of evidence-based practice to improve student outcomes. While the use of scientific evidence in education was introduced by NCLB in 2001, sixteen years later, there is little evidence to suggest that it is being done. Instruments to assess attitudes, barriers, and use of EBP have been developed in the fields of medicine, social work, and mental health. However, no comparable measures exist in education. The development of such an instrument could assist school districts in assessing needs and developing professional development activities to address those needs.

The purpose of this study was to examine the preliminary psychometric properties of the newly developed EBPSAT using a sample of teachers, administrators, and school-based related service providers. The EFA resulted in a two-factor structure, with 14 of the original 20 items being retained and the deletion of the *Practice* subscale. Expert analysis of the items of the EBPSAT indicated that the content experts found the majority of the items to be relevant. However, the CVI did not assess the comprehensiveness of the measure. Analysis of known groups validity indicated that the EBPSAT was able to differentiate between BCBAs, teachers,

and administrators. Finally, participants generally indicated that they found the instrument to be acceptable and that they found some value in the information collected by the measure.

Further research into the EBPSAT is necessary. The EFA suggested revisions to instrument, including the deletion of six items. Future research should investigate the wording of the items, specifically directing the questions to teachers. Researchers should also consider the addition of items to further address barriers to EBP implementation. Finally, future studies should include larger samples of teachers that are more representative of the general population.

References

AlGhamdi, K. M., Moussa, N. A., AlEssa, D. S., AlOthimeen, N., & AlSaud, A. S. (2013). Perceptions, attitudes, and practices toward research among senior medical students. *Saudi Pharmaceutical Journal*, *22*, 113-117. doi:10.1016/j.jsps.2013.02.006.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Asmussen, K. (2011). *The Evidence-Based Parenting Practitioner's Handbook.* Abingdon: Routledge.

Baker, S., Gersten, R., Dimino, J. A., & Griffiths, R. (2004). The Sustained use of research-based instructional practice. *Remedial & Special Education, 25*(1), 5-24.

Beauchat, K. A., Blamey, K. L., & Walpole, S. (2009). Building preschool children's language and literacy one storybook at a time. *Reading Teacher*, *63*(1), 26-39.

Bentler, P.M. & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78–117.

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the gold standard: Lessons from the history of RCTs. *New England Journal of Medicine*, *374*(22), 2175-2181. doi:10.1056/NEJMms1604593.

Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, *22*(4), 195-201. doi:10.4103/1117-1936.173959.

Brownlee, J., Sumsion, J., Irvine, S., Berthelsen, D., Farrell, A., Walsh, K., Ryan, S. & Mulhearn, G. (2015) Taking an evaluative stance to decision-making about professional

development options in early childhood education and care, *Early Years*, *35*(4), 411-426. doi:10.1080/09575146.2015.1099519.

Burns, M. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education*, 175-184. doi:10.1007/s10864-012-9158-9.

Burns M. K., Ysseldyke J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *Journal of Special Education*, *43*, 3–11.

Bussières, A. E., Terhorst, L., & Leach, M. (2015). Self-reported attitudes, skills and use of evidence-based practice among Canadian doctors of chiropractic: A national survey. *Journal of the Canadian Chiropractic Association, 59*(4), 332-348.

Cain, T. (2016) Research utilisation and the struggle for the teacher's soul: A narrative review. *European Journal of Teacher Education*, *39*(5), 616-629. doi: 10.1080/02619768.2016.1252912

Cartwright, N., & Hardy, J., (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press.

Castillo, J., March, A., Yin Tan, S.., Stockslager, K., & Brundage, A. (2016) Relationships between ongoing professional development and educators' beliefs relative to response to intervention. *Journal of Applied School Psychology, 32*(4), 287-312. doi: 10.1080/15377903.2016.1207736.

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *Journal of Special Education, 47*(2), 71-82. doi:10.1177/0022466911420877.

Cook, B. G., Cook, L., & Landrum, T. J. (2013). Moving research into practice: Can we make dissemination stick? *Exceptional Children*, *79*(2), 163-180.

Cook, D. A. (2014). When I say validity. *Medical Education, 48,* 948-949. doi:10.1111/medu.12401.

Craig, J. R., & Kacer, B. A. (2000). Using an innovation configuration component map to assess the relationship between student achievement and the degree of implementation of extended school services in a sample of Kentucky middle schools. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4) 349-354. doi: 10.1037/h0047358.

Danforth, S. (2006). From epistemology to democracy: Pragmatism and the reorientation of disability research. *Remedial & Special Education, 27*(6), 337-345.

Davis, L. (1992). Getting the most from your panel of experts. *Applied Nursing Research,* 5, 194-197.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Dedrick, R.F., & Greenbaum, P.E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders, 19* (1), 27-40. doi:10/1177/1063426610365879.

Detrich, R., Keyworth, R., & States, J. (2016). Leveraging evidence-based practices: From

policy to action. *Learning Disabilities*: *A Contemporary Journal*, *14*(2), 121-142.

Diaz, C. & Drewery, S. (2016) A critical assessment of evidence-based policy and practice in social work. *Journal of Evidence-Informed Social Work*, *13*(4), 425-431.

Drisko, J., & Grady, M. (2015). Evidence-based practice in social work: A contemporary perspective. *Clinical Social Work Journal, 43*(3), 274-282. doi:10.1007/s10615-015-0548-z.

Duffy, M., Giordano, V. A., Farrell, J. B., Paneque, O. M., & Crump, G. B. (2008). No Child Left Behind: Values and research issues in high-stakes assessments. *Counseling & Values*, *53*(1), 53-66.

Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency reliability. *British Journal of Psychology*, *105*, 399-412. doi:10.1111/bjop.12046.

Emmons, M., Keefe, E. B., Moore, V. M., Sánchez, R. M., Mals, M. M., & Neely, T. Y. (2009). Teaching information literacy skills to prepare teachers who can bridge the research-to-practice gap. *Reference & User Services Quarterly*, *49*(2), 140-150.

Fleischman, S., Scott, C., & Sargard, S. (August, 2016). Better evidence, better choices, better schools: State supports for evidence-based school improvement and the Every Student Succeeds Act. *Washington, DC: Center for American Progress.*

Frazier, S. L., Formoso, D., Birman, D., & Atkins, M. S. (2008). Closing the research to practice gap: Redefining feasibility. *Clinical Psychology: Science & Practice*, *15*(2), 125-129. doi:10.1111/j.1468-2850.2008.00120.x.

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage Publications.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007) *Educational research: An introduction*. Boston: Pearson Education.

Garland, R. (1991). The midpoint on a rating scale: Is it desirable? *Marketing Bulletin, 2*, 66–70.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Gore, J. M., & Gitlin, A. D. (2004). [Re] Visioning the academic-teacher divide: Power and knowledge in the educational community. *Teachers & Teaching, 10*(1), 35-58.

Gray, M., & Schubert, L. (2012). Sustainable social work: Modeling knowledge production, transfer, and evidence-based practice. *International Journal of Social Welfare*, *21*(2), 203-214. doi:10.1111/j.1468-2397.2011.00802.x.

Green, C. E., Chen, C. E., Helms, J. E., & Henze, K. T. (2011). Recent reliability reporting practices in psychological assessment: Recognizing the people behind the data. *Psychological Assessment*, *23*(3), 656-669. doi:10.1037/a0023089.

Gulamhussein, A. (2013). Teaching the teachers: Effective professional development in an era of high stakes accountability. *National School Boards Association Center for Public Education*. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Staffingstudents/Teaching-the-Teachers-Effective-Professional-Development-in-an-Era-of-High-Stakes-Accountability/Teaching-the-Teachers-Full-Report.pd

Guyatt, G. H. (1991). Evidence-based medicine [Editorial]. *American College of Physicians J Club*, *114*(2), A16. doi:10.7326/ACPJC-1991-114-2-A16.

Hahs-Vaughn, D.L. (2016). *Applied multivariate statistical concepts.* New York, NY; Routledge.

Hamby, T., Icke, W., & Babcock, M. (2016). Evidence for context switching in the effects of

average item length and item-length variability on internal consistency. *Journal of Personality Assessment, 98*(5), 491-502. doi:10.1080/00223891.2016.1147044.

Hanley, P., Chambers, B., & Haslam, J. (2016). Reassessing RCTs as the 'gold standard': Synergy not separatism in evaluation designs. *International Journal of Research & Method In Education, 39*(3), 287-298. doi:10.1080/1743727X.2016.1138457.

Hartley, J., & Betts, L. R. (2010). Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology*, *13*(1), 17-27. doi:10.1080/13645570802648077.

Haskins, R. (2015). Evidence speaks loudly at Results First. *Brookings.*

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): Construction of the schedule. *Journal of Psychology, 10*, 249–254.

Hattie, J. & Cooksey, R.W. (1984). Procedures for assessing the validities of tests using the "known-groups" method. *Applied Psychological Measurement*, *8*, 295-305. doi:10.1177/014662168400800306.

Helfrich, C. D., Li, Y.-F., Sharp, N. D., & Sales, A. E. (2009). Organizational readiness to change assessment (ORCA): Development of an instrument based on the Promoting Action on Research in Health Services (PARIHS) framework. *Implementation Science : IS*, *4*, 38. http://doi.org/10.1186/1748-5908-4-38.

Hitchcock, J., Kratchowill, T., & Chezan, L. (2015). What Works Clearinghouse standards and generalization of single-case design evidence. *Journal of Behavioral Education, 24*(4), 459-469. doi:10.1007/s10864-015-9224-1.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education.

Exceptional Children, *71*(2), 165-179.

Huat See, B., Gorard, S., & Siddiqui, N. (2016) Teachers' use of research evidence in practice: A pilot study of feedback to enhance learning, *Educational Research*, *58*(1), 56-72. doi: 10.1080/00131881.2015.1117798.

Jahan F, Maqbali, A.A., Siddiqui M.A., Zadjali, N.M.A. (2015) Attitude and barrier towards research amongst health care professionals working in primary care service of Oman. *Health Education Research and Development, 3*(144). doi:10.4172/2380-5439.1000144.

Javeri, M. & Persichitte, K. (2007). Use of Innovation Component Configuration Map (ICCM) to measure technology integration practices of higher education faculty. *International Journal of Technology in Teaching and Learning. 3,* 28-50.

Jones, M. L. (2009). A study of novice special educators' views of evidence-based practices." *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 32*(2), 101–120.

Juniel, P. M. (2015). Evidence-based practices: An exploratory study concerning school district professional development considerations. Retrieved from http://search.proquest.com.libproxy.nau.edu/docview/1767205697?accountid=12706

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Kourea, L., & Lo, Y. (2016). The educational validity and utility of single-case design research in building evidence-based practices in education. *International Journal of Research & Method in Education*, *39*(4), 349-364. doi:10.1080/1743727X.2016.1160278.

Landrum, T. J., Cook, B. G., Tankersley, M., & Fitzgerald, S. (2002). Teacher perceptions of the trustworthiness, usability, and accessibility of information from different sources. *Remedial & Special Education*, *23*(1), 42.

Leask, M. & Younie, S. (2013) National models for continuing professional development: The challenges of twenty-first-century knowledge management. *Professional Development in Education*, *39(*2), 273-287, doi: 10.1080/19415257.2012.749801.

Lehman, J., Greener, M., Simpson, D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, *22*(4), 2002, 197-209, doi: 10.1016/S0740-5472(02)00233-7.

Levin, B. (2004) Making research matter more. *Education Policy Analysis Archives* 12(56).

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637.

Maeda, H. (2015). Response option configuration of online administered Likert scales. International Journal of Social Research Methodology, *18*(1), 15-26. doi:10.1080/13645579.2014.885159.

Maheady, L., Rafferty, L. A., Patti, A. L., & Budin, S. E. (2016). Leveraging change: Influencing the implementation of evidence-based practice to improve outcomes for students with disabilities. *Learning Disabilities—A Contemporary Journal*, *14*(2), 109-120.

Martin (2013). Exploratory factor analysis [Class handout]. Department of Educational Psychology, Northern Arizona University, Flagstaff, AZ.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal

consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*(1), 28-50. doi:10.1177/1088868310366253.

Messick S. (1989). Validity. In: Linn RL, ed. Educational Measurement, 3rd edition. New York, NY: American. Council on Education, Macmillan 1989, 13–103.

Mitchell, R., 2013. What is professional development, how does it occur in individuals, and how may it be used by educational leaders and managers for the purpose of school improvement? *Professional Development in Education*, *39*(3), 387–400. 10.1080/ 19415257.2012.762721.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, 1*42*(2), 71-89. doi:10.1080/00221309.2014.994590.

Neal, J., Neal, Z., Kornbluh, M., Mills, K., & Lawlor, J. (2015). Brokering the research-practice gap: A typology. *American Journal of Community Psychology*, *56*(3/4), 422-435. doi:10.1007/s10464-015-9745-8.

Newman I, Lim J, Pineda F. (2013) Content validity using a mixed methods approach: Its application

and development through the use of a table of specifications methodology. *Journal of Mixed Methods Research*, 7(3), 243–60. doi: 10.1177/1558689813476922.

No Child Left Behind Act of 2001, Pub. L. 107-110, §115 Stat. 1425 (2001).

Noguera, P. A. (2009). The achievement gap: Public education in crisis. *New Labor Forum*, 18(2), 60-69. doi:10.4179/NLF.182.0000008.

Norman, G. (2004). Editorial: Beyond PBL. *Advances in Health Sciences Education*, 9(4),

257–260. doi:10.1007/s10459-004-5539-8.

Overview of No Child Left Behind. (2008). *Congressional Digest,* 87(5), 131-137.

Polit, D. & Beck, C. (2006). *Essentials of nursing research: Methods, appraisal, and utlilization.* Philadelphia. Lippincott Williams & Wilkins.

Procter, R. (2015). Teachers and school research practices: The gaps between the values and practices of teachers. *Journal of Education for Teaching, 41*(5), 464-477. doi:10.1080/02607476.2015.1105535.

Rey, O., & Gaussel, M. (2016). The conditions for the successful use of research results by teachers: Reflections on some innovations in France. *European Journal of Teacher Education, 39*(5), 577-587. doi:10.1080/02619768.2016.1260117.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research,* 27(2), 94.

Sackett, D. L., & Rosenberg, W. M. (1996). Evidence based medicine: What it is and what it isn't. *BMJ: British Medical Journal (International Edition),* 3*12*(7023), 71-72.

Sahin, I., & Yildirim, A. (2016). Transforming professional learning into practice. *ELT Journal: English Language Teaching Journal, 70*(3), 241-252. doi:10.1093/elt/ccv070.

Shea, C. M., Jacobs, S. R., Esserman, D. A., Bruce, K., & Weiner, B. J. (2014). Organizational readiness for implementing change: a psychometric assessment of a new measure. *Implementation Science, 9*(1), 1-35. doi:10.1186/1748-5908-9-7.

Sheldon, B. and Chilvers, R. (2002) An empirical study of the obstacles to evidence-based practice. *Social Work and Social Sciences Review, 10*(1), 6-26.

Sherman, W. (2008). No Child Left Behind: A legislative catalyst for superintendent action to

eliminate test-score gaps?. *Educational Policy*, *22*(5), 675-704.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher,* 31(7), 15–21.

Sparks, S. D. (2016). Finding Evidence-Based Fixes for Schools. *Education Week, 36*(6), 8-9.

Sparks, S.D. (2016). IES launches new version of research clearinghouse. *Education Week, 36*(5), 4.

Spencer, T. D., Detrich, R., & Slocum, T. A. (2012). Evidence-based practice: A framework for making effective decisions. *Education and Treatment of Children, 35*(2), 127-151.

Stevenson, M., Hedberg, J., O'Sullivan, K.A., & Howe, C. (2016) Leading learning: The role of school leaders in supporting continuous professional development. *Professional Development in Education, 42* (5), 818-835, doi:10.1080/19415257.2015.1114507.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics.* Boston: Pearson Education.

Terhorst, L., Leach, M., Bussières, A., Evans, R., & Schneider, M. J. (2016). Evaluating the psychometric properties of the evidence-based practice attitude and utilization survey. *Journal of Alternative & Complementary Medicine, 22*(4), 328-335. doi:10.1089/acm.2015.0221

Thompson, B. (2000). Q-technique factor analysis: One variation on the two-mode factor analysis of variables. In *Grimm,* L.G. & Yarnold, P., Reading and Understanding More Multivariate Statistics. Washington, D.C.: American Psychological Association.

Timperley, H. S., & Parr, J. M. (2007). Closing the achievement gap through evidence-based

inquiry at multiple levels of the education system. *Journal of Advanced Academics*, 19(1), 90-115.

Trybus, M. (2007). Understanding scientifically based research: A mandate or decision making tool?. *Delta Kappa Gamma Bulletin*, *73*(4), 5-8.

US Department of Education Institute of Education Sciences National Center for Educational Evaluation and Regional Assistance (2003). Educational practices supported by rigorous evidence: A user-friendly guide. Retrieved from https://bblearn.nau.edu/bbcswebdav/pid-2545078-dt-content-rid-.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2015), National Assessment of Educational Progress (NAEP), 2015 Mathematics and Reading Assessment. Retrieved from https://www.nationsreportcard. gov/reading_math_2 015/# mathematics?grade=4.

US Department of Education Office of Special Education and Rehabilitation Services (2010). Thirty-five years of progress in educating children with disabilities through IDEA. Retrieved from https://www2.ed.gov/about/offices/list/osers/idea35/history/ index_pg10.html.

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, *36*(2), 299-316. doi:10.1080/01411920902919257.

Vanlommel, K., Vanhoof, J., & Van Petegem, P., (2016) Data use by teachers: the impact of motivation, decision-making style, supportive relationships and reflective capacity, *Educational Studies*, *42*(1), 36-53, doi: 10.1080/03055698.2016.1148582.

Whitehurst, G. J. (2002). Evidence-based education. Paper presented at the Student

Achievement and School Accountability Conference. Retrieved from http://www2.ed.gov/nclb/methods/whatworks/eb/edlite-index.html.

Wike, T., Bledsoe, S., Manuel, J., Despard, M., Johnson, L., Bellamy, J., & Killian-Farrell, C. (2014). Evidence-based practice in social work: Challenges and opportunities for clinicians and organizations. *Clinical Social Work Journal*, *42*(2), 161-170. doi:10.1007/s10615-014-0492-3.

Williams, D., Lee, C. Harrison, C. and P. Black. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, *11*(1): 49–65.

Wrigley, T. (2016). Not so simple: The problem with 'evidence-based practice' and the EEF toolkit. *Comprehensive Education*, *58*(2), 237-252. doi:10.15730/forum.2016.58.2.237.

Yell, M. L., Shriner, J. G., & Katsiyannis, A. (2006a). Individuals with Disabilities Education Improvement Act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus On Exceptional Children*, *39*(1), 1-24.

Yell, M. L., Katsiyannas, A., & Shiner, J. G. (2006b). The No Child Left Behind Act, adequate yearly progress, and students with disabilities. *Teaching Exceptional Children*, *38*(4), 32-39.

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar,

A. (2015, June 1). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, *4*(2), 165-178. doi:10.15171/jcs.2015.017.

APPENDICES

Appendix A:

IRB Approval

NORTHERN ARIZONA UNIVERSITY

Institutional Review Board for the
Human Research Subjects Protection

128 South Knolls Dr
Building 56, suite 240
PO Box: 4087
Flagstaff AZ 86011
928-523-9551
928-523-1607 fax
http://nau.edu/Research/Compliance/Human-Subjects/Welcome

To:	Tiffany Sellars, B.A.
From:	NAU IRB Office
Approval Date:	May 5, 2017
Project:	The Evidence-Based Practice Self Assessment Tool: Evidence of Reliability and Validity
Project Number:	977615-1
Submission:	New Project
Review Level:	Exempt Review
Action:	EXEMPT
Project Status:	Exempt
Review Category/ies:	**Exempt Approval 45 CFR 46.101(b)(2):** Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior.

This submission meets the criteria for exemption under 45 CFR 46.101(b). This project has been reviewed and approved by an IRB Chair or designee.

Northern Arizona University maintains a Federalwide Assurance with the Office for Human Research Protections (FWA #00000357).

All research procedures should be conducted in full accordance with all applicable sections of the guidance.

Exempt projects do not have a continuing review requirement.

This project should be conducted in full accordance with all applicable sections of the guidance and you should notify the IRB immediately of any proposed changes that affect the protocol.

Amendments to exempt projects that change the nature of the project should be submitted to the Human Research Subjects Protection (HRSP) office for a new determination. See the guidance Exempt Research for more information on changes that affect the determination of exemption. Please contact the HRSP to consult on whether the proposed changes need further review.

You should report any unanticipated problems involving risks to the participants or others to the IRB.

All documents referenced in this submission have been reviewed and approved. Documents are filed with the HRSP Office. If subjects will be consented, the approved consent(s) are attached to the approval notification from the HRSP Office.

Appendix B:

Participant Informed Consent

Statement of Purpose: The purpose of this study is to determine the attitudes, opportunities and barriers to teachers' and related service providers' implementation of evidence-based practice. Survey participants are requested to complete an online survey that will take approximately 10 minutes.

Participation: Your participation in this survey is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty.

Risks and Benefits: You will receive no direct benefits from participating in this research study. However, your responses may help us learn more about the use of evidence-based practice in educational settings. There are no foreseeable risks involved in participating in this study other than those encountered in day-to-day life.

Confidentiality: Your survey answers will be sent to a link at SurveyMonkey.com where data will be stored in a password protected electronic format. Survey Monkey does not collect identifying information such as your name, email address, or IP address. Therefore, your responses will remain anonymous. No one will be able to identify you or your answers, and no one will know whether or not you participated in the study.

Compensation: You can also choose to participate in a prize drawing for one of two $50 Visa gift cards. You can enter into this drawing regardless of whether or not you participate in the online survey.

Contact: If you have questions at any time about the study or the procedures, you may contact the principal researcher, Tiffany Sellars via email at ts693@nau.edu or the faculty research advisor, Trina Spencer, Ph.D., via email at tds95@nau.edu. If you feel you have not been treated according to the descriptions in this form, or that your rights as a participant in research have not been honored during the course of this project, or you have any questions, concerns, or complaints that you wish to address to someone other than the investigator, you may contact the Northern Arizona University Internal Review Board at IRB@nau.edu.

ELECTRONIC CONSENT: By clicking the NEXT button you are verifying the following information

- You have read the above information
- You voluntarily agree to participate
- You are 18 years of age or older

Appendix C:

Letter to the Participants

Dear Teacher, Administrator, or Related Service Provider,

I invite you to participate in a very brief online survey that will be part of my doctoral dissertation. My work is related to the use of evidenced-based practice in education by administrators, teachers and others in achieving student outcomes.

Your participation would include completing a 10-minute online survey within the next three weeks. By completing this brief survey, you will be contributing to the development of a new instrument designed to help schools assess the extent to which they use research in practice. As a teacher, administrator, or related service provider, your input is extremely valuable.

If you choose to participate, simply click on the link to the survey provided below.

You can also choose to participate in a prize drawing for one of two $50 Visa gift cards by clicking on the link below and entering your email. You can enter into this drawing regardless of whether you participate in the online survey or not.

Thank You,
Tiffany Sellars
Ph.D. Candidate
College of Education
Northern Arizona University
ts693@nau.edu

Appendix D:

Expert Informed Consent

Statement of Purpose: The purpose of this study is to assess the content validity of the newly developed Evidence Based Practice Self-Assessment Tool. The survey will ask you about your expertise in evidence-based practice and about the relevance of items designed to assess attitudes, barriers, and use of evidence in practice. For the purpose of this survey, evidence-based practice is defined as is a decision making process that is informed by three sources: a) best available research evidence; b) clinical expertise; and c) client and family characteristics, values, and preferences. Importantly, in the items below the word "practice" refers to all professional activities of an individual and not a single intervention or program.

Participation: Your participation in this survey is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty.

Risks and Benefits: The benefits of this research to you are minimal except that you will have an opportunity to contribute to the development of a new instrument. The data from this study will be used as part of a dissertation project, for publications in academic journals and academic presentations. There are no expected risks associated with your participation in this study.

Confidentiality: Your survey answers will be sent to a link at SurveyMonkey.com where data will be stored in a password protected electronic format. Survey Monkey does not collect identifying information such as your name, email address, or IP address. Therefore, your responses will remain anonymous. No one will be able to identify you or your answers, and no one will know whether or not you participated in the study.

Contact: If you have questions at any time about the study or the procedures, you may contact the principal researcher, Tiffany Sellars via email at ts693@nau.edu or the faculty research advisor, Trina Spencer, Ph.D., via email at tds95@nau.edu. If you feel you have not been treated according to the descriptions in this form, or that your rights as a participant in research have not been honored during the course of this project, or you have any questions, concerns, or complaints that you wish to address to someone other than the investigator, you may contact the Northern Arizona University Internal Review Board at IRB@nau.edu.

ELECTRONIC CONSENT: By clicking the NEXT button you are verifying the following information

· You have read the above information
· You voluntarily agree to participate
· You are 18 years of age or older

Appendix E:

Letter to the Experts

Dear *****,

You have been identified as an expert in the area of evidence-based practice by my dissertation committee chair Dr. Trina Spencer. I invite you to participate in a very brief on-line survey that will be part of my doctoral dissertation.

Your participation would include completing a 10-minute on-line survey within the next three weeks. By completing this brief survey, you will be contributing to the development of a new instrument designed to help organizations assess the extent to which they use research in practice. Given your interest and expertise in this area, your input is extremely valuable.

If you choose to participate, simply click on the link to the survey provided below.

Thank You,
Tiffany Sellars
Ph.D. Candidate
College of Education
Northern Arizona University
ts693@nau.edu

Appendix F:

EBP Self-Assessment Tool

Evidence-Based Practice Self-Assessment Tool

Organization

Evidence-based Practice (EBP) is a decision making process that is informed by three sources: a) best available research evidence; b) clinical expertise; and c) client and family characteristics, values, and preferences. Importantly, in the items below the word "practice" refers to all professional activities of an individual and not a single intervention or program.

Please indicate the extent to which you agree with each statement by selecting the best option.

14. EBP is central to the mission and philosophy of my organization.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	I don't Know
○	○	○	○	○

15. Monitoring and sharing outcomes are essential in my organization's culture.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	I Don't Know
○	○	○	○	○

16. Fiscal resources are used to support the implementation of EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	I Don't Know
○	○	○	○	○

17. Administrators regularly supply resources (e.g., articles, reports, professional development) to support the implementation of EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	I Don't Know
○	○	○	○	○

18. EBP is implemented in my organization.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	I Don't Know
○	○	○	○	○

Evidence-Based Practice Self-Assessment Tool

Individual

Evidence-based Practice (EBP) is a decision making process that is informed by three sources: a) best available research evidence; b) clinical expertise; and c) client and family characteristics, values, and preferences. Importantly, in the items below the word "practice" refers to all professional activities of an individual and not a single intervention or program.

Please indicate the extent to which you agree with each statement by selecting the best option.

19. I value EBP as a decision making framework.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

20. I have sufficient reason or motivation to implement EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

21. I have sufficient resources and support to implement EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

22. I have sufficient time to implement EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

23. When making decisions, I value the characteristics, values, and preferences of my client(s).

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

24. When making decisions, I value clinical/field experience.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

25. When making decisions, I value research evidence.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

26. I regularly use professional resources other than research studies (e.g.,websites, practice guides) to inform my practice.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

27. I regularly use client data to inform my practice.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

28. I regularly use research findings to inform my practice.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

29. I have an ethical and professional responsibility to implement EBP.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

Evidence-Based Practice Self-Assessment Tool

Practice

Please indicate whether or not you have done each activity in the past eight weeks.

30. In the past eight weeks I have read a research study.

Yes	No
○	○

31. In the past eight weeks I have shared findings of a research study with a colleague, administrator, or client/family member.

Yes	No
○	○

32. In the past eight weeks I have shared progress data with a colleague, administrator, client/family member.

Yes	No
○	○

33. In the past eight weeks I have collected data to examine the impact of practice.

Yes	No
○	○

Appendix G:

Demographic Questionnaire

Evidence-Based Practice Self-Assessment Tool

Demographic Information

1. What is your age?

- ◯ 18-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45-54
- ◯ 55+

2. What is your gender?

- ◯ Female
- ◯ Male
- ◯ Other

3. Which race/ethnicity best describes you? (Please choose only one.)

- ◯ American Indian or Alaskan Native
- ◯ Asian / Pacific Islander
- ◯ Black or African American
- ◯ Hispanic
- ◯ White / Caucasian
- ◯ Multiple ethnicity / Other (please specify)

4. What is the primary language you speak?

- ◯ English
- ◯ Spanish
- ◯ Other

2

5. What is the highest degree or level of school you have completed?

- ◯ High School Diploma or Equivalent
- ◯ Some College, No Degree
- ◯ Associate's Degree
- ◯ Bachelor's Degree
- ◯ Master's Degree
- ◯ Professional Degree
- ◯ Doctoral Degree

6. In what year did you most recently graduate?

7. What is your title/position?

- ◯ Teacher's Aide
- ◯ Paraprofessional
- ◯ Teacher
- ◯ Administrator
- ◯ Intervention Specialist
- ◯ Speech Language Pathologist
- ◯ School Psychologist
- ◯ Occupational Therapist
- ◯ Other (please specify)

8. How long have you been in your current position?

- ◯ Less than 1 year
- ◯ 1-5 years
- ◯ 6-10 years
- ◯ 11-15 years
- ◯ 16-20 years
- ◯ More than 20 years

9. In what state or U.S. territory do you live?

10. Please describe the setting in which you work. (check all that apply)

- [ ] School
- [ ] Clinic Hospital
- [ ] Rural
- [ ] Suburban
- [ ] Urban

11. If you are a teacher, which grade(s) do you teach? (check all that apply)

- [ ] Not Applicable
- [ ] Kindergarten
- [ ] 1st grade
- [ ] 2nd grade
- [ ] 3rd grade
- [ ] 4th grade
- [ ] 5th grade
- [ ] 6th grade
- [ ] 7th grade
- [ ] 8th grade
- [ ] High School

12. If you are a teacher, which subject(s) do you teach? (check all that apply)

- [ ] Not Applicable
- [ ] English/Language Arts
- [ ] Math
- [ ] Science
- [ ] History
- [ ] Physical Education
- [ ] Music
- [ ] Other (please specify)

13. Did you take any research-based courses in college?

- ( ) Yes
- ( ) No

Please list by course name.

5

Appendix H:

Follow-Up Questions

Evidence-Based Practice Self-Assessment Tool

Follow-up Questions

The following questions ask your opinion about the previous sections of the survey (Organization, Individual, and Practice). Please indicate the extent to which you agree with each statement below. In the comment boxes please provide any additional feedback that could help improve this survey.

34. I think the data from this survey will be useful in my workplace.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

Comments:

35. I think the information collected by this survey is valuable.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

Comments:

10

36. I will change my practice after taking this survey.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

Comments:

37. The questions on this survey were relevant to my profession and easy to answer.

Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
○	○	○	○

Comments:

38. Do you think any question from this survey should be changed? If so, which ones/how?

39. Do you think any question should be added to this survey? If so, which one(s)?

APPENDIX I

Content Validity Questionnaire

Content Validity Questionnaire

Organization

The following questions are intended to assess EBP at the organization level. For the purpose of this survey, "organization level" refers to values, characteristics, and/or activities of an organization that promote or impede EBP implementation. For each of the questions below, please indicate the item's relevance to the content domain. Although you are are being asked to rate the relevance of each item, when respondents complete this survey, they will indicate the extent to which they agree with each item on a Likert scale from 1-5.

7. **How relevant is this item to the domain "EBP at the Organization Level"?**
EBP is central to the mission and philosophy of my organization.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

8. **How relevant is this item to the domain "EBP at the Organization Level"?**
Monitoring and sharing outcomes are essential in my organization's culture.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

9. **How relevant is this item to the domain "EBP at the Organization Level"?**
Fiscal resources are used to support the implementation of EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

4

10. **How relevant is this item to the domain "EBP at the Organization Level"?**
Administrators regularly supply resources (e.g., articles, reports, professional development) to support the implementation of EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

11. **How relevant is this item to the domain "EBP at the Organization Level"?**
EBP is implemented in my organization.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

5

Content Validity Questionnaire

Individual

The following questions are intended to assess EBP at the individual level. For the purpose of this survey, "individual level" refers to personal values, characteristics, and/or activities that promote or impede EBP implementation. For each of the questions below, please indicate the item's relevance to the content domain. Although you are are being asked to rate the relevance of each item, when respondents complete this survey, they will indicate the extent to which they agree with each item on a Likert scale from 1-5.

12. **How relevant is this item to the domain "EBP at the Individual Level"?**
I value EBP as a decision making framework.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

13. **How relevant is this item to the domain "EBP at the Individual Level"?**
I have sufficient reason or motivation to implement EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

14. **How relevant is this item to the domain "EBP at the Individual Level"?**
I have sufficient resources and support to implement EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

15. **How relevant is this item to the domain "EBP at the Individual Level"?**
I have sufficient time to implement EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

16. **How relevant is this item to the domain "EBP at the Individual Level"?**
When making decisions, I value the characteristics, values, and preferences of my client(s).

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

17. **How relevant is this item to the domain "EBP at the Individual Level"?**
When making decisions, I value clinical/field experience.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

18. **How relevant is this item to the domain "EBP at the Individual Level"?**
When making decisions, I value research evidence.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

19. **How relevant is this item to the domain "EBP at the Individual Level"?**
I regularly use professional resources other than research studies (e.g.,websites, practice guides) to inform my practice.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

20. **How relevant is this item to the domain "EBP at the Individual Level"?**
I regularly use client data to inform my practice.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

21. **How relevant is this item to the domain "EBP at the Individual Level"?**
I regularly use research findings to inform my practice.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

22. **How relevant is this item to the domain "EBP at the Individual Level"?**
I have an ethical and professional responsibility to implement EBP.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

Content Validity Questionnaire

Practice

The following questions are intended to assess the use of EBP in practice. For the purpose of this survey, "in practice" refers to professional activities that align with the definition of EBP. For each of the questions below, please indicate the item's relevance to the content domain. Although you are are being asked to rate the relevance of each item, when respondents complete this survey, they will indicate whether they have or have not completed each activity within the past eight weeks.

23. **How relevant is this item to the domain "EBP in Practice"?**
In the past eight weeks I have read a research study.

No Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

24. **How relevant is this item to the domain "EBP in Practice"?**
In the past eight weeks I have shared findings of a research study with a colleague, administrator, or client/family member.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

25. **How relevant is this item to the domain "EBP in Practice"?**
In the past eight weeks I have shared progress data with a colleague, administrator, client/family member.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

26. **How relevant is this item to the domain "EBP in Practice"?**

In the past eight weeks I have collected data to examine the impact of practice.

Not Relevant	Somewhat Relevant	Relevant	Very Relevant
○	○	○	○

Comments

10