Clinical Education Outcomes and Research Directions in Speech-Language Pathology: A Scoping Review

George W. Wolford  
*Midwestern University – Glendale*, gwolfo@midwestern.edu

Schea Fissel Brannick  
*Midwestern University*, sfisse@midwestern.edu

Sarah Strother  
*Midwestern University*, sstrother29@midwestern.edu

See next page for additional authors

DOI: https://doi.org/10.30707/TLCSD5.2.1624983591.656565
Clinical Education Outcomes and Research Directions in Speech-Language Pathology: A Scoping Review

Cover Page Footnote
Author Note George Wolford, PhD, MS CCC-SLP https://orcid.org/0000-0001-9318-5814 Schea N. Fissel Brannick, PhD, CCC-SLP https://orcid.org/0000-0003-1220-4193 Sarah Strother, BS Laura L. Wolford, PhD, CCC-SLP https://orcid.org/0000-0002-2823-0989 Laura L. Wolford, PhD, CCC-SLP, is currently employed as a copy-editor for TLCSD. At the time of the development and methods for this project, she was not yet employed by TLCSD. George W. Wolford, Schea Fissel Brannick, & Laura L. Wolford are employed by and draw a salary from Midwestern University. Sarah Strother was a graduate student at Midwestern University at the time of this project’s development. Correspondence concerning this article should be addressed to George Wolford, 5815 West Utopia, Glendale, AZ 85308. Email: gwolfo@midwestern.edu

Authors
George W. Wolford, Schea Fissel Brannick, Sarah Strother, and Laura Wolford

This scholarship of teaching and learning research is available in Teaching and Learning in Communication Sciences & Disorders: https://ir.library.illinoisstate.edu/tlcsd/vol5/iss2/3
Introduction

The early literature on supervision and clinical education in speech-language pathology (SLP) dates back to the 1960s (Anderson, 1988; Dudding et al., 2017) but has developed slowly since (Dudding et al., 2017; Shapiro, 1985). Rigorous research on clinical education emerged in the early-to-mid 1980s, partially in response to a position statement by the American Speech-Language-Hearing Association (ASHA), which asserted that the field had limited knowledge of the effectiveness of “critical factors in supervision methodology” (ASHA, 1978, p. 480). Current research on clinical education continues to be “sparse” (Dudding et al., 2017, p. 167), especially regarding best practices in clinical education. Often, recommendations for best practices draw on evidence from other fields (e.g., Anderson, 1988).

Within the SLP clinical education literature, Hagler et al. (1997) described publications in three categories: adulation, prescriptive, or descriptive. Adulation publications assert the importance of clinical education, while prescriptive publications describe clinical education theories and provide recommendations (e.g., Geller & Foley, 2009; Mawdsley & Scudder, 1989). Adulation and prescriptive works represent a substantial proportion of the clinical education literature, but neither type directly tests hypotheses. The prescriptive literature often presents methodological descriptions of how a clinical educator should facilitate student growth as part of a complex multifaceted model but does not test that model (e.g., Anderson, 1988; Cogan, 1973; Geller & Foley, 2009; Pickering et al., 1992). Descriptive studies, in contrast, are those that report an outcome or measured effect of a clinical supervisory practice or experience. Descriptive studies are the most valuable to determining best practices because they test hypotheses about an educational practice or method. Descriptive studies have empirical data. These studies form the basis of research for evidence-based education (Ginsberg et al., 2012), which allows clinical educators to follow best practices in education based on empirical knowledge.

Clinical education models based on prescriptive research are often complex, multifaceted, and place substantial demands on clinical educators “…to be skilled practitioners as well as effective teachers” (Higgs & McAllister, 2007, p. 187). When professional organizations translate these intricate models into recommendations, the result is a dauntingly long checklist of skills and responsibilities. For example, the Council of Academic Programs in Communication Sciences and Disorders ([CAPCSD]; 2013) identified that ASHA (2008) recommended 125 items either as knowledge or skills that a clinical educator should possess. When the CAPCSD (2013) attempted to categorize these 125 items as introductory, intermediate, or advanced skills, only 20 items were suggested to be advanced. This indicates that entry-level clinical educators, such as first-time clinical faculty members or student internship supervisors, are expected to demonstrate 105 skills or knowledge areas. With such a long list of competencies, assessing or tracking clinical educator practices becomes daunting.

Despite the many recommendations for clinical educators provided by prescriptive works, several recommendations are not based on rigorous descriptive study within the field that uses empirical data to justify its importance. As an example, a recent tutorial (Dalessio [Procaccini], 2019) endorsed the educational practice of strategic questioning based on evidence from other fields. The article included the caveat that more research about the practice is needed in the field of SLP because “Much of the current available research […] has been conducted in clinical fields outside.
speech-language pathology” (p. 1471). Providing a recommendation for further research within the field and borrowing findings from other fields is indicative of the presence of gaps in the SLP clinical education literature. Rigorous scholarship of teaching and learning (SoTL) research can generate principles that apply across disciplines (e.g., McKinney, 2013; Meyer & Land, 2003) and can be considered a valid source of evidence, especially with limited discipline-specific studies. However, a lack of discipline-specific studies to translate more general theories leads to two problems. The first problem is that SoTL is highly context-dependent and “inquiry varies by place, time, stakeholder, and sub-discipline” (Ginsberg et al., 2017, p. 1). Drawing evidence almost exclusively outside of the field removes the contextual factors and limits the strength of the evidence. The second problem goes to the process of searching for evidence in evidence-based education frameworks (Brown & Williams, 2015; Ginsberg et al., 2011). In describing searching for evidence in an evidence-based education (EBE) framework, McAllister (2015) describes the “student or student group” (p. 175) as a key component of determining whether the evidence applies. Attempting to apply an educational methodology studied with one population of students to another population is analogous to applying an intervention intended for one clinical population to different clinical population. Though the clinician may find similar outcomes between groups, best practice dictates that the efficacy of that clinical intervention should be evaluated with the population at hand. Likewise, if there were more discipline-specific SoTL research, one would be advised to examine that evidence for educational practices within the field.

Several possible reasons exist for the lack of evidence to back discipline specific educational practices. It may be in part because SoTL research historically has not been valued by higher education institutions (Ginsberg & Bernstein, 2011). For instance, dissertations related to clinical education may not lead to publication (e.g. Larson, 1981; Nilsen, 1983; Turner, 1994), and there is minimal grant funding available for SoTL research (Marquis, 2015). Additionally, the prescriptive models themselves may be difficult to test. Often a model includes complex and abstract recommendations, such as “self-supervision” (Anderson, 1988; McCrea & Brassuer, 2019), critical thinking (CAPCSD, 2013), or “self-reflection” (Schön, 1983). Abstract constructs reported in clinical education literature frequently lack consistent theoretical definitions (Caty et al., 2015) and are difficult to operationalize and measure. Clinical education constructs may therefore be interpreted and operationalized very differently by different professionals (Li et al., 2009). Validated tools that measure clearly defined constructs are required to quantify or describe student outcomes. However, prior reports of the clinical education literature indicate that such measures are not often used. As Shapiro (1985) writes, “much of the supervisory literature has focused on factors which were perceived by the participants to be effective […] rather than on the demonstration of actual change in the behavior of clinicians as a result of the supervisory process. Investigating whether supervisees do anything differently as a result of having met with the supervisors seems to be a minimal criterion for supervisory effectiveness” (p. 96).

Yet within the past decade, organizations and researchers nationally and internationally have shown renewed interest in the clinical education of graduate students in SLP. For example, in the United States, ASHA published revised standards for the training of clinical supervisors (ASHA, 2020). In Australia and other countries, researchers work to operationalize and describe the competencies required of graduate SLP students (Ho & McAllister, 2018; McAllister et al., 2011). Also, an increasing number of recent publications explore new modes of clinical supervision,
including the application of technologies such as simulated experiences (Benadom & Potter, 2011; Ferguson & Estis, 2018; Moineau et al., 2018).

Given the recent activity, it is timely to summarize the body of research on clinical education in SLP to describe what researchers are investigating and how that is measured. Recent reviews of clinical education research are narrow in scope, having focused on a particular construct (Caty et al., 2015) or being comprised of studies from other disciplines with few or no studies from SLP (Kühne et al., 2019; Milne et al., 2011). There is a need to broadly review and summarize the state of the literature on clinical education in the field to inform supervisory standards based on the present evidence.

**Aims.** This review aims to describe how researchers are investigating clinical education in speech-language pathology and how they are measuring learning outcomes. As a secondary aim, this document may serve as a roadmap for published outcome measures within the field to aid future researchers in developing their study.

**Methods**

**Design.** Given that the aim of this paper is to describe the activity within the field, which has been previously been reported as sparse (Dudding et al., 2017), a scoping review (Arksey & O’Malley, 2005) was determined to be the most appropriate method for reviewing the current body of work. A scoping review describes the literature and current activity by providing an overview of existing research in the field rather than guiding the reader to the best available evidence to answer a narrow question (Daudt, et al., 2013; Pham et al., 2014). Scoping reviews are considered useful as an initial summary of research activity in new/emerging fields or in fields where literature is very limited.

Arksey & O’Malley (2005) identified a six-step process to conducting a scoping review, which was followed for this review. These steps include, (a) identifying a broad research question; (b) searching for relevant studies; (c) determining the studies with inclusion and exclusion criteria; (d) “charting the data” (p. 22) according to key issues and themes; (e) “collating, summarizing and reporting” (p. 22) the data; and (f) optionally consult with key stakeholders. This final optional step was not applied for this review. A quality appraisal of studies that met inclusion criteria (during step c) was also conducted to increase the rigor of this scoping review in line with more recent methodological recommendations (Daudt et al., 2013).

**Step 1: Broad Research Question.** Two broad research questions guided the present scoping review. These questions aimed to evaluate SLP clinical education research that explores the effectiveness of supervision interventions using measurable student outcomes:

1. What are the aims of published clinical education research in speech-language pathology?
2. How is student learning being measured to represent outcomes of clinical education?

**Step 2: Search Procedures.** Our broad search criteria included published articles between 1970 and 2018 that reported at least one measurable student outcome. Specific search terms and strings were applied that related to clinical education and speech-language pathology into four electronic
databases by December of 2018 (CINAHL Plus with Full-Text, Medline Complete, ERIC, and PsycINFO) by the first and third authors. Full search terms can be found in the supplementary document (Appendix A). Two additional journals containing relevant SLP clinical education literature (Perspectives of the ASHA Special Interest Groups [Perspectives]¹; Teaching and Learning in Communication Sciences & Disorders [TLCSD]) were added via hand searches after noticing that their contents were mostly omitted from the electronic database searches. The initial search strategy yielded 1733 results. Additional hand searches of references cited in included articles published in 2013 or later were also conducted to reduce electronic search bias.

Step 3: Inclusion and Exclusion Criteria. After the initial search, abstracts were appraised based on the inclusion and exclusion criteria found in Table 1. The aim was to include descriptive articles that reported at least one measured student outcome variable within SLP, but exclude adulation or prescriptive papers (Hagler et al., 1997). If an article’s abstract was unavailable or unclear, the article was included at this stage.

Table 1
Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A described student learning or perceptual outcome (qualitative or quantitative)</td>
<td>Outcomes related to only supervisors and not students</td>
</tr>
<tr>
<td>At least one SLP student participant who was not identified as an undergraduate²</td>
<td>Outcomes from undergraduates or Clinical Educators (CEs) only²</td>
</tr>
<tr>
<td>Reference to active clinical practice or perceptions related to clinical practice by the student</td>
<td>Classroom practice not related to clinical education</td>
</tr>
<tr>
<td>A level of rigor described below in the full appraisal procedure</td>
<td>Article was not in English</td>
</tr>
<tr>
<td></td>
<td>Search resulted in non-article (i.e. conference proceedings or dissertation)</td>
</tr>
<tr>
<td></td>
<td>Article was published prior to 1970 or after 2018</td>
</tr>
</tbody>
</table>

¹ Of note, the Perspectives of the ASHA Special Interest Groups was recognized as a peer-reviewed journal in 2017 though “many articles prior to that publish date did undergo peer review” (ASHA Journals Academy, n.d.); publications in ASHA SIG journals prior to 2017 were included given that the journal was a major avenue of publication for authors reporting on clinical education. The Perspectives journals were searched by hand primarily since they did not appear to be fully catalogued by the other databases, unlike the other ASHA journals.

²Deciding how to treat the undergraduate data was a difficult decision. The aim was to try to include as many potential studies with students earning terminal clinical degrees; some countries allow for undergraduates to obtain full practice licensure while others do not. Inclusion criteria of “at least one SLP student not identified as an undergraduate” and also exclusion criteria of “outcomes from undergraduates or CEs only” meant that a study with mixed graduate and undergraduate populations would be included. Studies with only undergraduates would be excluded. Studies where the students were not described as undergraduates would be included. The reasoning behind this decision was that studies that included mixed populations would likely be aimed at describing students in pre-professional practice whereas an undergraduate only study may include research of students who are being educated for the next step of graduate school rather than for professional practice. It is unclear if undergraduate and graduate students are truly different “student groups” (McAllister, 2015), which may be an important step for future research.
Following abstract appraisal and removal of duplicates, 360 articles remained. All 360 articles were subjected to a full appraisal, which involved application of detailed inclusion/exclusion criteria and criteria for study quality (outlined in Figure 1). Upon review, a set of articles was discovered that reported student outcomes without a clear measurement system or tool, such as anecdotal data that students liked a clinical placement. For instance, Bedore et al. (2008) provided a detailed description of a program for bilingual trainees and only summarized a non-empirical outcome of training by stating that, “Students’ comments have generally been positive” (p. 271). The primary aim of such publications was to describe a program as a recommendation or suggestion for areas of future research, which aligns with prescriptive publications, rather than to test hypotheses related to student outcomes. The discovery of such articles without clearly measured outcomes led us to apply quality indicators (see Table 2) to better obtain the most relevant data that met the aims of this review.
Figure 1
Search Strategy and Appraisal Procedures

Articles through database search, n = 1733

- Records after abstract appraisal, n = 448
- Records after duplicates removed, n = 360

Full Appraisal Procedure

Inclusion
- A described outcome
- SLP Student who was not an undergraduate
- Clinical education or clinical outcome

Exclusion
- Undergraduate only
- No student outcomes (no supervisor only outcomes)
- Article not in English
- Non-article (e.g., dissertation)
- Undetected duplicate

Exclude

Quality Appraisal

Meets 2/5
- Sample size > 30
- Sample from more than 1 location
- Quantitative measure used or referenced qualitative methodology
- Use of statistics beyond descriptive statistics or systematic presentation of qualitative results
- Response rate > 50% OR controlled for outliers in any way

Meets 3/3
- Single-site research
- More than just general feedback
- Described organizational factors or situates experience

Count after Full Appraisal Procedure, n = 103

Final Corpus
n = 124

Articles identified in references from studies within the past 5 years, n = 21
Quality Appraisal. The primary quality indicator included any article that reported a measurement of student outcomes other than or in addition to student self-report. This criterion was incorporated due to historical concerns about the utility of self-report measures found within the field and concerns about the utility of self-report measures in other fields (Eva et al., 2004; Eva & Regehr, 2005, 2011; Shapiro, 1985). If a study met this primary indicator, it was retained. For studies containing only self-report measures, secondary criteria were applied based on quality measures adapted from Protogerou and Hagger (2019). Of these five secondary criteria (see Table 2), if at least two were met, the study was included. If two of those criteria were not met, tertiary criteria were applied. These tertiary criteria allowed us to retain several studies that described case descriptions of a program, term, or successful student. These studies may be important to several models that describe the situated experience of a student within an educational framework (e.g., Collins et al., 1988). If a study did not meet defined criteria at any level, it was excluded.

Table 2
Quality Appraisal Criteria

<table>
<thead>
<tr>
<th>Criteria Needed for Inclusion by Stage</th>
<th>Primary Requirement: must meet 1/1</th>
<th>Secondary Requirement: must meet 2/5</th>
<th>Tertiary Requirement: must meet 3/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Had a measure of student learning or growth outside of a self-report, quantitative scale (e.g., Likert scale of confidence), or self-reported open-ended perceptual response (e.g., focus groups, open-ended feedback questions on survey measures)</td>
<td>(1) A sample size that was either justified or at least 30 students.</td>
<td>(1) The article described a single-site research project</td>
<td></td>
</tr>
<tr>
<td>(2) The sample was taken from more than one location (e.g., two research sites, a national survey as opposed to a survey only of students within one university)</td>
<td>(3) A quantitative measure in the literature or referenced the specific qualitative methodology used to complete the analysis</td>
<td>(2) The article had more analysis than just general feedback (e.g., clear, logical, systematic presentation of qualitative themes, clear and logical presentation of data of descriptive statistics)</td>
<td></td>
</tr>
<tr>
<td>(3) They described the organizational characteristics or situated experience of students (e.g., types of teaching used, particular setup or protocol being implemented)</td>
<td>(4) The use of statistical analysis beyond descriptive statistics or clear, logical, systematic presentation of qualitative results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) A response rate over 50% or controlled for potential outliers in any way</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Following the three-tiered quality appraisal, 103 articles met inclusion criteria. Hand searches of articles from 2013 until 2018 yielded an additional 21 articles, which resulted in 124 articles included for final review (see Appendix B in the supplementary document for a full list of articles included in the corpus).

**Step 4: Identifying Key Dimensions.** Article data was extracted and categorized by each research question dimension, including the study purpose, supervision interventions, and student outcome measures to chart the broad topology of clinical education research in SLP.

**Coding Key Dimensions.** The data extracted were largely qualitative. A system of coding was needed to appropriately account for responses to our research questions. This coding system is described below.

**Purpose statement.** To characterize the qualitative descriptions of study purpose, dimensions of these qualitative statements were “organized thematically” (Arksey & O’Malley, 2005, p. 28) and coded using a bottom-up approach. This approach was adapted from the content analysis procedures presented by Baker et al. (2018) who proposed that researchers might avoid missing critical elements on a predetermined checklist by using a bottom-up approach to extract, summarize, and classify meaningful units of data based on the research reviewed. In this study, the goals driving each clinical education study were summarized by extracting the explicit statement of purpose, statement of aims, or stated research questions. If the study did not contain a stated purpose, aim, or research question, the first author selected the statement within the text that best described the study focus. The meaningful elements were extracted for analysis from all statements of purpose to develop qualitative themes. The first author developed the initial coding system, which was provided to the second author for feedback and then revised for consistency. Part of the revision process included examining themes identified from only one study, searching for similar themes to combine, and then discarding themes only found in one study. The second author combined the existing themes into broader themes which were discussed with the first author for revision. After themes were finalized, similar themes were further grouped into broadly connected categories to streamline discussions about similar constructs. These broad categories were termed “clusters” to minimize terminological confusion.

In developing themes and clusters, the authors attempted to create logical and theoretically consistent groupings. However, differences in terminology across countries, time, and theoretical orientations within the individual studies necessitated several subjective decisions within the coding. The results presented here are one of many potential interpretations of the data that the authors believe appropriately provides an overview of the breadth of the corpus. Much as Sandback et al. (2020) describes a systematic review as being “limited by the quality of evidence which they summarize” (p. 4), a scoping review’s map of the evidence is limited by the presentation and uniformity of descriptions contained within the studies. We address this limitation by presenting the operational definitions for the themes in the supplemental document (Appendix C) to increase methodological transparency and allow for replication of these coding methods.

**Student outcome measures.** Finally, the methods that researchers used to measure student outcomes were identified by extracting details of the measurement tool and the level of data generated. To be coded as an outcome, a measure must have been reported in the results section.
of the paper in addition to the methods. Student outcome measures were grouped by type after identifying similarities within the corpus. The supplementary document presents the operational definitions for the categories of outcome measures (Appendix D).

Results

The following results relate to step 5 of Arksey and O’Malley’s (2005) framework for collating and describing results.

Description of the Corpus. A total of 124 research articles met inclusion criteria. Although the corpus extends back to 1970, studies after 2010 comprise 72/124 (58%) of the data set. A graph of the year of publication is shown in Figure 2 below.

Figure 2
Articles by Year of Publication

Most studies were conducted in the United States [n = 91]; some were conducted in Australia and/or New Zealand [n = 17]. Research was also represented from Canada [n = 5], Hong Kong [n = 3], the United Kingdom [n = 3], Israel [n = 2], the Philippines [n = 2], Ireland [n = 1], Malaysia [n = 1], and Mexico [n = 1].

Fifty-one of the 124 studies met the primary quality appraisal criteria, indicating they used at least one measure that was not self-report. Forty-five studies met secondary criteria, and 28 studies met tertiary criteria. The majority of the corpus would have been excluded using only primary criteria.

Note two articles (Cook et al., 2019; Harmon et al., 2019) met inclusion criteria in 2017 with advance online publications. They are included in the 2016-2018 bracket in Figure 2 as they were considered a 2017 publication at the time of analysis though the citations throughout have been updated to reflect their final publication date.
**Research Purposes.** Following the methodological process for identifying and combining the meaningful elements into similar themes, 47 total themes were identified. Of those 47, 31 themes were found in fewer than five papers within the corpus. These themes are referred to as “low-frequency themes” throughout. Themes found in 5 or more studies, are referred to as “high-frequency themes.” This division is somewhat arbitrary though gives some indication of more and less frequent lines of inquiry. Of note, several of the high-frequency themes involved logical groupings of related, but inconsistent terms used across research reports. For instance, the studies that included a “student perceptions” theme described these perceptions in a variety of ways, such as “student perceptions,” “needs,” “student beliefs,” and “student preferences” (Alborés et al., 2017; Chipchase et al., 2012; Plexico et al., 2017; Wagner & Hess, 1997). Additionally, though many themes were connected by a common feature, not all studies containing that theme were intricately related. For instance, studies with a theme of “unique populations” identified the unique training needs that are required to work with a specific population. However, the specific population itself varied across studies, including clients with autism spectrum disorder (ASD) (Donaldson, 2015), fluency needs (Cardell & Hill, 2013), or cleft palate (Pamplona et al., 2015).

**Clusters.** These themes naturally fell into one of four groupings, termed “clusters” for the purposes of discussion: Outcome Measures (methods of assessment, growth, competency, predictor, simulation) \( n = 61 \), Student Perspectives (student perceptions, cognitive-emotional states, supervisory needs) \( n = 55 \), Teaching Methods (reflection, feedback, simulation to train, systems, self-evaluation) \( n = 48 \), and Educational Contexts (IPE/IPP, unique experience, unique populations) \( n = 45 \). The cluster is specified by theme in Appendix C. Several studies contained themes from more than one cluster.

The first cluster, Outcome Measures, was related to assessment of student performance. The top three themes of this cluster were “investigating different methods of assessment” \( n = 16 \), “student growth” \( n = 14 \), and “investigating competency” \( n = 12 \). Nine studies investigated predictors of student success in graduate school (e.g., Oratio & Hood, 1977), which usually focused on admissions metrics (e.g., Baggs et al., 2015; Reed, 2007). “Efficacy of simulations” was the purpose of 7 studies (e.g., Syder, 1996; Zriack et al., 2003), which consistently yielded positive results. Many \( n = 30 \) of the 61 studies contained low-frequency themes, and 15 of the 61 studies included only low-frequency themes. These low-frequency themes primarily related to development or description of specific student learning constructs, such as critical thinking (e.g., Miles et al., 2016), cultural competence (de Diego-Lázaro, 2018), or self-supervision (Donnelly & Glaser, 1993). Some were also more generic but did not neatly combine with any other themes from the initial coding, such as clinical skills (Messersmith & Brouwer, 2012) or discussion of remedial students (Means, 2005).

The second cluster, Student Perspectives, included investigations of student opinions or perceptions. The most common theme in the cluster was “assessing student perceptions of clinical education” \( n = 40 \). This was also the most common theme within the entire corpus. Researchers linked student perceptions to what students report they like or do not like in a supervisor (e.g., Atkins, 1996; Fencel & Mead, 2017), how much they feel they learned or their beliefs in their ability to complete tasks (e.g., Oswald, 2013; Pasupathy & Bogschutz, 2013), and what their opinions were following an (often unique) experience (Dowling, 1987; Opina-Tan, 2013). The terminology used to describe student perceptions was highly varied and included the terms,
“student perceptions,” “needs,” “student beliefs,” and “student preferences” (Alborés et al., 2017; Chipchase et al., 2012; Plexico et al., 2017; Wagner & Hess, 1997). When compared with others’ perceptions (e.g., clinical educators, standardized patients, clients), SLP student perceptions agreed (e.g., Carlin et al., 2012; Gerlach & Subramanian, 2018) and also differed (Gerlach & Subramanian, 2018; Mandel, 2015; Moineau et al., 2018; Rudolf et al., 1983; Smith & Anderson, 1982). The second most common theme was “cognitive-emotional states” \( [n = 7] \), where researchers investigated constructs such as student anxiety (Hill et al., 2013; Plexico et al., 2017; Sleight, 1985) or motivation (Ho & Whitehill, 2009) as they related to variables of interest such as simulation (Hill et al., 2013) or amount of time in graduate school (Plexico et al., 2017).

The third cluster, Teaching Methods \([n = 49]\), was related to the ways that clinical educators facilitated growth in students. The most significant finding within this cluster was the breadth of strategies used, but the limited depth of investigation for any given theme. Of the 49 studies included, the highest frequency themes were “reflection” \([n = 9]\), “simulation training” \([n = 7]\), and “feedback” \([n = 6]\). Eighteen of the 49 studies included only low-frequency themes, which investigated specific teaching interventions like “planning” (e.g., Peaper, 1984), “supervisory conferences” (e.g. Smith & Anderson, 1982), or “analysis of video-recordings” (e.g., Smith, 2015). Overall, the themes within the teaching methods cluster revealed surface-level investigations of methods, rather than an in-depth body of work about a given teaching method.

The fourth cluster, Educational Contexts, included studies that sought to describe the learning environment of the student, including specifying the types of clients with whom the students learned to work. The top themes within this cluster were “unique populations” \([n = 21]\), “interprofessional education/interprofessional practice (IPE/IPP)” \([n = 17]\), and “unique experiences” \([n = 9]\). The theme of “unique populations” was highly heterogenous and described many different client populations. Authors described the need to train students to work with a unique, demanding, or complex population, like clients with ASD (Donaldson, 2015), fluency needs (Cardell & Hill, 2013), or cleft palate (Pamplona et al., 2015). The studies of IPE/IPP typically included student perceptions of clinical education (e.g., Chipchase et al., 2012; Guitard et al., 2010; Opina-Tan, 2013; Renschler et al., 2016), building interprofessional teams (e.g., Cox et al., 1999; Peña & Quinn, 2003), or interprofessional learning (e.g., Harmon et al., 2019; Howell et al., 2011). Most interprofessional practice/interprofessional education studies \([n = 17]\) described one example of a successful interprofessional experience, rather than presenting a systematic exploration of features that contributed to positive interprofessional practice patterns or students’ acquisition of these skills. There were notably fewer low-frequency themes in this cluster including only “medical settings” (e.g., Warner et al., 2018), “e-supervision” (Carlin et al., 2013), and “externship” (Plexico et al., 2017, p. 7).

**Regularity within Lines of Research.** Publications of a given theme were also typically spread out in time. Once a theme was first written about, it was not the case that the theme would be given extensive attention, thoroughly investigated, and then considered established. Instead, the clinical education research appears to be sporadically spaced in time. Frequently, a study would venture into new directions and often identify itself as a pilot study or preliminary in nature (e.g. Cox et al., 1999; Hansen et al., 2017; Towson et al., 2018) or revisits decades-old research in a modern context (e.g., Plexico et al., 2017). The first and second authors of a study often initiated a line of
research but then did not appear to address it again. Histograms of the research over time (Figure 3) are presented for the top themes within that cluster as an illustration of this trend.

Figure 3

*Histograms of Themes Over Time*
Summary. The results of qualitative analysis of study purpose statements revealed substantial variation in the subjects that authors investigated, and many studies were of a preliminary nature. Purpose statements most frequently involved assessment of student performance, student perspectives, teaching methods, and educational contexts. Within the clusters, the distribution of themes was not uniform. Some clusters, such as Student Perceptions, were dominated by one high-frequency theme, but others, such as Teaching Methods, contained numerous low-frequency themes. The terminology used to refer to study constructs was inconsistent, and the lines of research were spread out in time.

Outcome Analysis. To answer the second research question, “how is student learning being measured to represent outcomes of clinical education?”, student outcome measures were grouped into categories (Figure 4). These categories included self-report scales, open-ended responses in questionnaires, behavioral observation scales, competency-based assessments, written content analysis, knowledge tests, and other measures that didn’t clearly fit within any category. The main finding in this area is that student self-reported outcomes are by far the most frequent measure employed across studies. In addition, within a category, any given measure is used infrequently. Studies often reported more than one measure. A description of the categories follows.

Figure 4
Frequency of Outcome Type

Self-Report Measures. The two most frequent categories of outcome measures involved student self-reports through scales and student responses to open-ended prompts, which were qualitatively
analyzed. These can be grouped conceptually as “self-report measures.” In total, 73 of 124 studies relied on a self-report measure as the only outcome measure for their research. The 51 remaining studies that included another measure also used self-report measures occasionally. This trend towards relying on self-report measures as the only outcome measure is not limited to early studies; it has continued throughout recent years as well (see Figure 5).

**Figure 5**
*Prevalence of Studies using Only Self-Report Measures Over the Years*

![Prevalence of Studies using Only Self-Report Measures Over the Years](image)

**Self-report Scales.** Self-report scales were the most frequently observed measure within the corpus. Of the 75 studies within the corpus that used self-report scales, most authors developed a unique scale for their study \( n = 56/75 \), rather than relying on one already established in the literature \( n = 19/75 \). The 19 articles that borrowed a scale from prior literature often cited a source other than a scholarly journal, such as a conference (e.g., Gouvier et al., 1979 as cited in Rudolf et al., 1983). Many of these studies did not report reliability or validity indices, though the psychometric properties of these scales were not evaluated as part of this review. Table 3 presents studies that used a previously published rating scale. Note that although different measurement scales were used, many studies attempted to measure similar constructs. No scale was used in more than three studies within the corpus.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Original/Citation1</th>
<th>Studies citing this article</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Supervisory Working Alliance Inventory: Trainee Form (SWAI)</td>
<td>Efstation et al. (1990)</td>
<td>Cassidy (2013)</td>
</tr>
<tr>
<td>A Self-Efficacy Scale: Conducting Therapy with Stuttering Adults and Children</td>
<td>Gouvier et al. (1979)</td>
<td>Rudolf et al. (1983)</td>
</tr>
<tr>
<td>The Sleight Clinician Anxiety Test (PRS)</td>
<td>Sleight (1985)</td>
<td>Plexico et al. (2017),</td>
</tr>
<tr>
<td>The Carolina Opinions on Care of Older Adults (COCOA)</td>
<td>Hollar et al. (2011)</td>
<td>McManus et al. (2017)</td>
</tr>
</tbody>
</table>

1Citations within this column should be considered secondary sources
The self-report scales summarized in Table 3 can be categorized broadly into five groups: the supervisee’s attitude towards interprofessional education/practice (Latrobe Community Health Service & the Health & Socialcare Interprofessional Network, 2009; Renschler et al., 2016), what they value or think they learned from the experience (e.g., de Diego-Lazáro, 2016), their wants or internal attitudes (e.g., Larson, 1981; Sleight, 1985; Tiien, 1983), their opinions of a population or workplace (McManus et al., 2017), and their appraisal of the supervisor (Efstation et al., 1990; Goodyear & Heppner, 1984). Researchers who did not create their own measures often used student outcome measures that were not discipline-specific or were developed to measure perceptions from a variety of disciplines (e.g., Efstation et al., 1990; McFadyen et al., 2007).

**Open-ended responses.** Researchers in the field used a variety of methodologies to collect responses to open-ended prompts, including questionnaires [n = 17], interviews [n = 10], focus groups [n = 6], multiple methods [n = 4], letters to themselves [n = 1], photo-elicitation interviews [n = 1], and one was unclear [n = 1]. Four studies used multiple methods. For some papers, the open-ended responses were the primary means to answer the central research question (e.g., Croker et al., 2015; Opina-Tan, 2013), while other studies used them more as a supplementary outcome. For instance, Bressman and Eriks-Brophy (2012) used open-ended questions to expand upon closed-set answers from other survey measures.

A variety of data analysis methods were employed to evaluate open ended responses. The most common data analysis method was described as either thematic analysis or analysis of themes [n = 15] followed by, grounded theory [n = 4], phenomenological analysis with various subtypes [n = 4], a hermeneutic approach [n = 2], and phenomenological hermeneutics [n = 2]. Two authors referenced general analysis frameworks such as inductive analysis (Gustafsson et al., 2016). Thirteen studies did not provide a clear data analysis framework or theory used to report and summarize results.

Of note, open-ended measures have become more prevalent in the field over time, as shown in Figure 6 below, suggesting a growing acceptance of qualitative methodology.
**Behavioral Observational Scales.** Behavioral observation scales were used to measure targeted aspects of particular clinical skills that students performed, rather than more generic sets of competencies (see Appendix E for a full list of measures). A specific behavioral observation scale was rarely repeated across studies, which is consistent with other categories of outcomes within the corpus. In each scale, the number of behaviors measured ranged substantially. For instance, Weltsch & Crowe (2006) picked one behavior to work on for each of three students, while Kaplan & Dreyer (1974) operationally defined and measured eight interpersonal verbal behaviors, seven nonverbal behaviors, and nine speech-directed behaviors. Two studies used a fidelity checklist as an outcome measure (Donaldson, 2015; Lorio et al., 2016).

The outcomes from behavioral observation scales led to conclusions that students perform better after training compared to baseline performance (e.g., Towson et al., 2018, Weltsch & Crowe, 2006). Results also suggest that students make progress over time when data is examined cross-sectionally; more advanced students typically demonstrate more sophisticated skills (e.g., Moses & Shapiro, 1996). Studies that reported a baseline design with no intervention or partial training showed that student behavior changes once supervision intervention is implemented (Gillam et al., 1990; Herd, 2009).
**Competency-Based Assessments.** Competency-based assessments either explicitly stated that they were measuring competency or measured generic clinical skills rather than specific clinical behaviors. For instance, one of the 5-point ratings in Duthie and Robbins (2013) was “student clinician provides sufficient models for producing target objectives” (p. 11). This rating reflects a perceived ability to model the target rather than a measurement the number of models that the student provided; the latter would be a behavioral observation scale. The specific measures are shown in Table 4 below.

Table 4

<table>
<thead>
<tr>
<th>Competency-based assessment</th>
<th>Number of studies</th>
<th>Studies using this assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPASS® (McAllister et al., 2013) (or portions of thereof)</td>
<td>3</td>
<td>Hill et al. (2014), Ho et al. (2014), Ho &amp; McAllister (2018)</td>
</tr>
<tr>
<td>Clinical Fellowship Skills Inventory (ASHA, n.d.)</td>
<td>1</td>
<td>Baggs et al. (2015)</td>
</tr>
<tr>
<td>Assessment of Foundation Clinical Skills (AFCS)</td>
<td>1</td>
<td>Hill et al. (2014)</td>
</tr>
<tr>
<td>The UWO Clinical Grading System</td>
<td>1</td>
<td>Johnson &amp; Shewan (1998)</td>
</tr>
<tr>
<td>18 variable system (Oratio, 1976)</td>
<td>1</td>
<td>Oratio &amp; Hood (1977)</td>
</tr>
<tr>
<td>Clinical Skill Acquisition Rubric (CSAR) (Resnick, et al., 2014)</td>
<td>1</td>
<td>Troche &amp; Towson (2018)</td>
</tr>
<tr>
<td>Graduate Student Development Profile (GSDP)</td>
<td>1</td>
<td>Hancock &amp; Brundage (2010)</td>
</tr>
<tr>
<td>Clinical Practicum Competency Checklist</td>
<td>1</td>
<td>Kjelgaard &amp; Guarino (2012)</td>
</tr>
<tr>
<td>Departmental Knowledge and Skills (KASA) form</td>
<td>1</td>
<td>Hart et al. (2008)</td>
</tr>
<tr>
<td>WPACC (Shriberg et al., 1974)</td>
<td>1</td>
<td>Shriberg et al. (1977)</td>
</tr>
</tbody>
</table>

While many studies did not provide the validation data or clear guidelines about their competency-based assessment (e.g., Duthie & Robbins, 2013; Oratio & Hood, 1977; Wagner & Malandraki, 2016), several studies developed (Johnson & Shewan, 1998) or added aspects to existing competency based-assessments (Hill et al., 2014). Some of these competency-based assessments refer to departmental or unpublished works (e.g., Peaper, 1988 as cited in Perkins & Mercaitis,
In contrast to the majority of investigations in this category, a few studies did perform some degree of validation for a competency-based assessment (e.g., Johnson & Shewan, 1988). COMPASS® (McAllister et al., 2013) was the most frequently referenced measure within the data set and has undergone extensive validation (McAllister et al., 2011).

**Written Content Analysis.** Written content analysis was defined as an assessment of the accuracy or content within students’ written work. In written content analysis, there is an additional step of coding, determining accuracy, or counting the frequency of an objective construct. Studies that used written content analysis fell into one of two groups: (a) analysis of students’ written reflections or (b) analysis of written answers to key content questions relative to a scenario or structured observation. Analyses included measurements such as number of words (Donnelly & Glaser, 1993), type of reflection (Cluver, 1988), or depth and breadth of reflection (Cook et al., 2019). The methods of coding reflections were not replicated across studies, with the exception of counting the number of words (Donnelly & Glaser, 1993; Meilijson & Katzenberger, 2009). However, methodology in Cook et al. (2019) drew from prior work with physiotherapy and undergraduate SLP students (Hill et al., 2012). While not all of the analyses of written reflections found significant results (Cluver, 1988), most studies noted a positive change in the reflective practice of the participants for at least one outcome coded (Cook et al., 2019; Donnelly & Glaser, 1993; Meilijson & Katzenberger, 2009;).

Studies that used written student outcome measures to examine answers to content questions consistently used unique criteria that was only established within that study. These sorts of analyses were often linked to assessing knowledge of best practices or quality care. For instance, Ferguson & Estis (2018) examined six specific critical elements related to pre-term infant feeding as part of a simulation. None of these measures were replicated within the corpus.

**Other, Grades, and Knowledge Tests.** Outcomes within the “other” category were coded as such because they did not fit neatly into any of the above common categories \( n = 8 \). These studies are summarized in the supplementary document (see Appendix F). Some studies within this category were similar. Two studies used a think aloud or talk aloud method (Boyer, 2013; Ginsberg et al., 2016), two studies used simulation as a method to gain unique measures such as agreement with experts or speed of task completion (Dudding & Pfeiffer, 2018; Strang & Meyers, 1987), and two studies used a methodology that fell somewhere in between written content analysis and open-ended responses (Peña & Quinn, 2003; Wainscott, 2016).

Seven studies used grades as an outcome measure. These included clinical grades, summative GPA, and a pass/fail or a multitiered pass/fail system (Ho & Whitehill, 2009). No study used grades as the only outcome measure. Of the four studies that used a knowledge test, three created their own tests on the information being taught in the study (Ferguson & Estis, 2018; Horton et al., 2004; Wilson, Chasson, et al., 2017). One study used a previously published measure: the Autism Knowledge Survey-Revised (AKS-R; Swiezy, et al., 2005; Wilson, Chasson, et al., 2017).

**Summary.** Researchers have measured student learning with a wide range of different outcome categories consisting mostly of varied self-report measures (either self-report scale or an open-ended response), as well as behavioral observations or competency-based measures. Most outcome measures across categories were unique to one study and were not replicated across studies.
Authors have a variety of potential outcome measures to use or adapt to their needs, but none were commonly used to measure any construct apart from COMPASS® (McAllister et al., 2013) to measure competency.

**Discussion**

This scoping review sought to map the broad state of the clinical education literature in SLP by summarizing two areas of the literature: the purposes of the studies that have been conducted and how those outcomes were measured. Results of this review identified a wide breadth of investigations that used outcome measures mostly unique to each study. Though there are generally positive outcomes along the breadth of constructs investigated, these constructs were largely measured by self-report scales with little replication. However, the body of work provides a variety of emerging evidence that clinical education is effective for teaching students how to work with a wide range of populations and provides illustrations of successful learning situations. Additionally, this review found that clinical education publications within the field are on the rise within the past ten years, which speaks to a growing interest towards deepening the body of knowledge regarding clinical education in speech-language pathology.

**Self-Report Measures and Alternatives.** Consistent with previous observations (Shapiro, 1985), this review found that student outcomes continue to be largely measured by self-report. The most frequent theme within the corpus investigated student perceptions, which has resulted in a large body of literature describing how students view given situations. The most frequent outcome category was student self-report scales followed by responses to open-ended questions. Student self-reported data is valuable to answer research questions about student perceptions, internal states, and to provide a starting point to future research. In addition, self-report measures can be easily generated by researchers and quickly completed by student participants. However, the finding that many studies used self-report measures to answer research questions that are unrelated to student perceptions is cause for concern. One of the lines of research within the corpus involved instructing students to develop self-analysis, self-reflection, and observation skills—ostensibly because these are skills that students need to hone. Using these same untrained skills to provide the only outcome measure for a study seems questionable, and research from other fields indicates that student self-report and self-assessment measures are not reliable reflections of performance (Eva et al., 2004; Eva & Regehr, 2005, 2011). Future studies should attempt to include at least one direct measure not based on student self-report.

Despite the widespread use of student self-report as an outcome measure, there are many studies that used other measures and therefore provide more compelling evidence. Several of these outcome measures are flexible enough that they could be adapted for a variety of learning outcomes. For instance, Shapiro and Anderson (1989) had students and supervisors agree upon certain commitments during the supervisory conference, which included a wide variety of tasks. They then measured the completion of those commitments. While the framework is broad, flexible, and yielded significant results between conditions, no other subsequent study within this review used the same outcome.

A variety of other outcome measures were found within the corpus and are available to future researchers for replication. Several researchers have: developed specific measurements for key
behaviors (e.g., timing measures for endoscopy from Benadom & Potter, 2011), used competency-based assessments such as COMPASS® (McAllister et al., 2013), and systematically analyzed the quality of student reflections (Cook et al., 2019). Only a few studies used grades as an outcome measure \( [n = 7] \). If grades are a true measure of student performance or learning, it is surprising that more studies did not use grades as outcome measures.

Measuring specific student behaviors during simulated experiences is a unique and potentially promising methodology. Through simulations, a researcher can control the clinical scenario and allow multiple students to engage in the same experience. By controlling the clinical scenario, researchers can assess observable behavioral outcomes in concert with other measures such as agreement with experts (Dudding & Pfeiffer, 2018), timing within simulation (Benadom & Potter, 2011), skill demonstration (Towson et al., 2018), and standardized assessment administration (Moineau et al., 2018).

Though researchers seeking to measure SLP students’ learning tended to start from scratch in each study by developing a new tool, they may be better served by replicating or adapting measures from prior work for future studies. Given the current variability in outcome measures, comparing the performance of students from study to study is made more complicated in an area of research that is already complex with many potential variables. Furthermore, validated outcome measures would give researchers better tools to conduct randomized control trials with large numbers of participants, which would yield more compelling evidence (e.g. Hill et al., 2020).

**Isolated Lines of Inquiry.** Although this review spans 50 years, it is still difficult to clearly delineate the lines of research in clinical education. The descriptive research mapped out by this scoping review does not neatly correspond with several theoretical models in the field (e.g., Anderson, 1988; Collins et al., 1988; Geller & Foley, 2009). Instead, many studies are isolated, describing a unique educational context or student views of a unique question. This isolation is compounded by high variability in definitions of constructs and use of outcome measures. One difficulty in providing generalizable educational recommendations is that outcome measures are often unique to each study, even when the purpose of many investigations are similar. This makes it difficult to draw clear parallels between studies’ outcomes.

An unexpected finding within the corpus was the emergence of the Educational Context cluster, in which the themes highlighted the unique aspects of students’ educational environment. Since no educational outcome has been deeply researched within the field, it was anticipated that key components of well-known prescriptive models would have been investigated in regularly occurring clinical contexts, rather than in less frequent, unique educational contexts. In fact, the opposite occurred. One reason may be that these studies were designed to showcase how a master clinical educator or highly efficient program trains their students. However, the unique features make translating practice patterns unclear for the day-to-day practice of a clinical educator who does not often encounter the same circumstances.

Some of these unique studies were rigorous and met primary appraisal criteria yet were disconnected from the rest of the corpus. For instance, Pamplona et al. (2015) investigated the role of mentorship in addition to ongoing supervision in a cleft palate clinic using a randomized control trial design. Within the corpus, it was the only study that compared an ongoing typical supervision
group with a supervision group that had additional mentorship, the only study from a cleft palate clinic, and the only study from Mexico. Despite the rigor of the work, the study does not fit clearly within the corpus. One would expect, for instance, that this paper might be situated in a line of research with prior studies regarding the role of mentorship or further studies regarding mentorship in cleft palate clinics, but as yet, it stands alone. The lack of connection or follow-up from these high-quality studies is somewhat surprising. Since the uniqueness of these studies means they stand distinct from more typical clinical education experiences, it is also unclear if the experiences they describe demand different sets of behaviors from students or clinical educators.

**Best Practices in Clinical Education.** Within the cluster of Teaching Methods, which focused on investigating the efficacy of concrete clinical educational practices, there were a wide range of low-frequency themes. Rather than in-depth investigation of a few teaching methods, this review found that the corpus contained many teaching methods that underwent only surface-level investigations. In addition, there is little research that focuses on which of two clinical education techniques is more effective or what makes a particular clinical education practice lead to positive student learning outcomes. The recently implemented requirements for SLP clinical educators to complete continuing education in clinical education (ASHA, 2020) highlight an important and timely need for more systematic investigations of clinical teaching methods that lead to measurable student outcomes, on which best practice recommendations can be based. The present lack of in-depth studies of clinical educator teaching methods within the field limits the voracity of the continuing education instruction and threatens the validity of best practice recommendations for SLP clinical educators.

This finding may indicate that SLPs continue to draw from other fields to develop their educational practices, since teaching methods within the field have not been investigated in depth (Anderson, 1988; Dudding et al., 2017). Borrowing from other bases of educational theory and research is not uncommon within healthcare fields (O’Brien & Battista, 2020; Teunissen, 2010). However, as O’Brien and Battista (2020) point out, if care is not taken to be rigorous in understanding the scope of an educational theory and to ensure that these concepts fit well within the field, researchers “run the risk of misappropriating theory if scholars lack awareness of or misconceive the purpose, paradigmatic stance, scope, limitations, and terminology” (O’Brien & Battista, 2020, p. 484). While practice patterns drawn from other fields may be valuable, how well they apply to SLP is largely unstudied. The SLP scope of practice is wide, and its educational trajectory is substantially different from fields such as nursing or medical education, which have deeper bases of educational literature specific to their fields. It is not clear that all literature on education translates equally well into an SLP clinical education context.

**Recommendation for future research**

Based on the discussion above, the following recommendations would address gaps in the literature base:

1. Researchers should continue the trend of investigating clinical education more frequently to generate a larger corpus of research especially with a focus on best practices. Substantial gaps within the literature exist such that the exploration of nearly any supervisory teaching technique to make recommendations for best practices would benefit the field.
2. There is a pressing need to define and validate outcome measures related to student learning. Constructs should be theoretically justified and operationally defined with reference to previous research.

3. Researchers should seek to measure student learning constructs using direct measurement of learning behaviors in addition to self-reported indices.

4. Future investigations interested in student perceptions should focus on application of validated perceptual measures, validating perceptual measures, or correlation of self-reported measures with observable behaviors.

5. Promising supervision interventions from other fields, which have little to no current evidence within this corpus, should be investigated to establish their efficacy within the field of SLP.

6. Numerous promising outcomes from unique educational contexts (e.g., Pamplona et al., 2015) should be investigated further to determine if similar practices facilitate student growth in other settings and with other populations.

7. Researchers should explore previously validated measurement systems drawn from allied disciplines such as clinical psychology and physical therapy to determine if these measurement tools are applicable within SLP.

Limitations

While every effort was taken to ensure a quality and transparent review, this review was not without limitations. While a number of checks were put into place such as the use of multiple databases and hand searches, the search strategy did not include all databases and likely did not uncover all potential published works since 1970. In addition, the work here is the interpretation of the authors, which is subject to bias. The frequency of studies from the United States was high and may indicate that the search strategy was biased towards the authors’ home country. Decisions such as exclusion of undergraduate-only studies, search terms, and hand searches of United States-based journals may also have contributed to bias. In particular, it is unclear whether studies that focused solely on undergraduate students should have been included. Educational requirements for SLPs are different from country to country and have changed substantially during the time period studied in the current paper. Future research should consider evaluating studies that focused on undergraduate populations. In addition, while the authors determined that using the purpose statements from an article would be the best way to reduce their own bias, a poorly worded or especially concise statement of purpose within a publication would limit the number of coded themes. A final limitation is that the authors did not register a protocol at the start of the study, which could have been used to reduced bias.

Conclusion

This scoping review of 124 publications is the most comprehensive map of the state of the clinical education in SLP research to date. While many studies previously have noted a sparsity of research (Dudding et al., 2017), this review has described the gaps found in the literature base. The wide scope of the literature in clinical education is mostly exploratory in nature and needs more in-depth exploration of best practices by using measures outside of student self-report. It is hoped that this scoping review will provide future researchers with a picture of the current gaps in the literature and a means of finding prior work upon which they might build. The appendices attached are as
detailed as possible to provide that roadmap for a corpus which is challenging to navigate. Several studies have promising designs, initial findings, and outcome measures that authors can build upon in developing new research.

Acknowledgements

The authors would like to thank the library staff for fulfilling countless Interlibrary Loan requests throughout the duration of this project. This manuscript would not have been possible without their support.

Author Disclosures

Laura L. Wolford, PhD, CCC-SLP, is currently serves as a copy-editor for TLCSD. At the time of the development and methods for this project, she was not yet providing service to TLCSD.

George W. Wolford, Schea Fissel Brannick, & Laura L. Wolford are employed by and draw a salary from Midwestern University. Sarah Strother was a graduate student at Midwestern University at the time of this project’s development.

References

References are marked with an asterisk if they were included in this scoping review corpus. For a full list, see Appendix B.


* Cluver, L. P. (1988). Journal writing by speech-language pathology student clinicians. *The Clinical Supervisor, 6*(1), 101–107. [https://doi.org/10.1300/j001v06n01_08](https://doi.org/10.1300/j001v06n01_08)


---

4 As referenced in Cassidy (2013).


* Lorio, C. M., Delehanty, A. D., & Woods, J. J. (2016). Digital platforms and supervisory feedback to graduate student clinicians. *Perspectives of the ASHA Special Interest Groups, J*(11), 18–34. [https://doi.org/10.1044/persp1.sig11.18](https://doi.org/10.1044/persp1.sig11.18)
* Mandel, S. (2015). Exploring the differences in expectations between supervisors and supervisees during the initial clinical experience. *Perspectives on Administration and Supervision*, 25(1), 4-30. [https://doi.org/10.1044/aas25.1.4](https://doi.org/10.1044/aas25.1.4)


* Oswalt, J. (2013). Students’ self-perceptions of knowledge and skills pre and post externship. Perspectives on Administration and Supervision, 23(3), 110-120. https://doi.org/10.1044/aas23.3.110


Peaper, R. E. (1984). An analysis of student perceptions of the supervisory conference and student developed agendas for that conference. *The Clinical Supervisor, 2*(1), 55–69. [https://doi.org/10.1300/j001v02n01_06](https://doi.org/10.1300/j001v02n01_06)


presentation]. American Speech-Language-Hearing Association Annual Convention, Orlando, FL, United States.


* Smith, H. C. (2015). Student clinicians self-evaluating performance by reviewing session videos and using a rubric. *Perspectives on Administration and Supervision, 25*(2), 78-84. [https://doi.org/10.1044/aas25.2.78](https://doi.org/10.1044/aas25.2.78)


* Troche, J., & Towson, J. (2018). Evaluating a metric to predict the academic and clinical success of masters students in speech-language pathology. *Teaching and Learning in Communication Sciences & Disorders, 2*(2), 1-31. [https://doi.org/10.30707/tlcsd2.2troche](https://doi.org/10.30707/tlcsd2.2troche)


* Wainscott, S. (2016). Building collaborative competencies for speech language pathologists and deaf educators: Review of a pilot project. *Perspectives of the ASHA Special Interest Groups, 1*(10), 31–42. [https://doi.org/10.1044/persp1.SIG10.31](https://doi.org/10.1044/persp1.SIG10.31)


## Appendix A

### Search Terms by Database

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Terms</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline Complete</td>
<td>Database</td>
<td>“speech language pathology” or “speech pathology” or “speech* therapy” or “communication disorder*” or “communication science*”) AND (“clinical education” or “clinical educator” or supervision or supervisor OR supervisee) ”</td>
<td>Expanders off, default search box, 2nd search 11/2019</td>
</tr>
<tr>
<td>CINAHL, PsycINFO, and ERIC</td>
<td>Database</td>
<td>(“speech language pathology” or “speech pathology” or “speech* therapy” or “communication disorder*” OR “communication science*”) AND (“clinical education” or “clinical educator” or “supervision” or “supervisor” OR “supervisee”)</td>
<td>Expanders off, default search box, 2nd search 11/2019</td>
</tr>
<tr>
<td>Perspectives of the ASHA Special Interest Groups</td>
<td>Journals</td>
<td>“supervision OR &quot;clinical education&quot; OR supervisor OR supervisee” was used and only “articles”</td>
<td>Only articles</td>
</tr>
<tr>
<td>Teaching and Learning in Communication Sciences &amp; Disorders</td>
<td>Journal</td>
<td>All articles searched by hand</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Full List of Articles Included in the Corpus


What is the impact of interprofessional supervision on speech-language pathology student learning outcomes?


Daly, G. (2010). Supervision: Formative assessment as a clinical supervision tool. Perspectives on Administration and Supervision, 20(3), 113-116. https://doi.org/10.1044/aas20.3.113


Fitzgerald, M. D. T. (2009). Reflections on student perceptions of supervisory needs in clinical education. *Perspectives on Administration and Supervision, 19*(3), 96-105. [https://doi.org/10.1044/aas19.3.96](https://doi.org/10.1044/aas19.3.96)

Fredrickson, T., & Moore, S. (2014). Key factors of influence in clinical educator relationships. *Perspectives on Administration and Supervision, 24*(1), 12-20. [https://doi.org/10.1044/aas24.1.12](https://doi.org/10.1044/aas24.1.12)


Lorio, C. M., Delehanty, A. D., & Woods, J. J. (2016). Digital platforms and supervisory feedback to graduate student clinicians. Perspectives of the ASHA Special Interest Groups, I(11), 18–34. https://doi.org/10.1044/persp1.sig11.18

Mandel, S. (2015). Exploring the differences in expectations between supervisors and supervisees during the initial clinical experience. Perspectives on Administration and Supervision, 25(1), 4-30. https://doi.org/10.1044/aas25.1.4


### Appendix C

**Operational Definitions of Themes & Cluster for Stated Purposes (n = 124)**

<table>
<thead>
<tr>
<th>Purpose Themes</th>
<th>Cluster</th>
<th>Description of Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Perceptions</td>
<td>Student Perspectives</td>
<td>Reference to attempting to understand student perceptions or perspectives of an experience or event</td>
</tr>
<tr>
<td>Unique Populations</td>
<td>Educational Contexts</td>
<td>Explicit mention to a clinical population as a defining element of the purpose (e.g., Wilson, Chasson, et al., 2017)</td>
</tr>
<tr>
<td>IPE/IPP</td>
<td>Educational Contexts</td>
<td>Reference to interprofessional practice or interprofessional education as a key element</td>
</tr>
<tr>
<td>Methods of Assessment</td>
<td>Outcome Measures</td>
<td>Specific to the design attempting to determine if an assessment methodology “fits” or discussion of a new assessment method; simple labeling of the methods of assessment used did not qualify</td>
</tr>
<tr>
<td>Growth</td>
<td>Outcome Measures</td>
<td>Mention of student growth over time; also studies seeking to describe students at different point in their development to understand differences between more novice and more advanced students</td>
</tr>
<tr>
<td>Competency</td>
<td>Outcome Measures</td>
<td>Mentions the word competency, aims to delineate growth in student competency, or discusses competency</td>
</tr>
<tr>
<td>Self-Reflection</td>
<td>Teaching Methods</td>
<td>Mention of self-reflection, reflection, reflective practice, or journal reflections; this reference must include a focus on reflection or the teaching of reflection and not just use as an outcome.</td>
</tr>
<tr>
<td>Unique Experiences</td>
<td>Educational Contexts</td>
<td>Explicit reference to characterizing the situation or experience as special, different, unique</td>
</tr>
<tr>
<td>Predictor</td>
<td>Outcome Measures</td>
<td>Aims to predict student performance based on certain metrics; several of these include admissions prediction that link application variables with student clinical performance later (e.g., Baggs et al., 2015).</td>
</tr>
<tr>
<td>Cognitive-Emotional States</td>
<td>Student Perspectives</td>
<td>Reference to study of linking cognitive emotional states of students (e.g., anxiety, empathy, motivation) to student outcomes</td>
</tr>
<tr>
<td>Feedback</td>
<td>Teaching Methods</td>
<td>Explicitly mentioned the word feedback</td>
</tr>
<tr>
<td>Simulation Training</td>
<td>Teaching Methods</td>
<td>Reference to using a simulation to train clinical skills or problem-solving</td>
</tr>
<tr>
<td>Simulation Efficacy</td>
<td>Outcome Measures</td>
<td>Reference to studying if simulation was a viable method of assessing or teaching</td>
</tr>
<tr>
<td>Systems</td>
<td>Teaching Methods</td>
<td>Some reference to changing a system or overall set of procedures related to day-to-day practice (not a one-time training) to determine impact of systemic variables on student performance (e.g., Banigan &amp; Cronen [1988] examined impacts of changing forms)</td>
</tr>
<tr>
<td>Self-evaluation</td>
<td>Teaching Methods</td>
<td>Referenced student self-evaluation</td>
</tr>
<tr>
<td>Supervisory Needs</td>
<td>Student Perspectives</td>
<td>References examining the supervisory needs of a student</td>
</tr>
</tbody>
</table>

*Note: This data is only for the high frequency themes. Definitions of low-frequency themes are available by request to the first author. The themes were developed using a bottom-up framework and then combined and confirmed using these definitions. Using these or other definitions a priori may have yielded different results.*
### Appendix D

**Outcome Measures by Type**

<table>
<thead>
<tr>
<th>Term</th>
<th>n</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Report Scale</td>
<td>75</td>
<td>Numeric scale (usually a likert-style scale around 5 choices) detailing student perceptions of an experience. Others (e.g., SPs, CEs, parents) may have completed these as well.</td>
</tr>
<tr>
<td>Open-ended responses</td>
<td>39</td>
<td>Free response in any form (e.g., interviews, written items in surveys, focus groups, phone call interviews) related to perceptions about an experience.</td>
</tr>
<tr>
<td>Behavioral Observation Scale</td>
<td>17</td>
<td>Scales that defined specific observable behaviors and used frequency counts or % of occurrence. These scales had discrete criteria (e.g., implemented behavior vs did not implement behavior, counts of certain behaviors) and were not defined as addressing a competency. Fidelity checklists were grouped into this category if they included discrete observations.</td>
</tr>
<tr>
<td>Competency-based assessment</td>
<td>15</td>
<td>CE ratings of student competencies. Most often these were identified explicitly as rating “competencies” rather than behaviors or skills. In the absence of such a distinction, the scale was identified as a competency-based rating scale rather than a behavioral rating scale only if the scale or rubric score specified (a) general acquisition of skills or abilities over time (b) rather than performance within a specific session. These scales were never student self-report only.</td>
</tr>
<tr>
<td>Other Measures</td>
<td>8</td>
<td>Other measures were a heterogenous grouping of measures that do not fit neatly into any framework.</td>
</tr>
<tr>
<td>Grades</td>
<td>7</td>
<td>Specific assigning of a reported grade in a class. This could be a letter grade, %, pass/fail demarcation, or GPA. Most often found in studies geared toward predicting success in a given program or clinic possible for admissions.</td>
</tr>
<tr>
<td>Written Content Analysis</td>
<td>7</td>
<td>Denotes an analysis of the content of the written work. Written content analysis must have some degree of analysis of “accuracy” or “quality” of the measure. For instance, these might be written responses to a scenario that are judged for accuracy or quality. This would not include taking the perceptions of an experience as the outcome data, which would be open-ended questions.</td>
</tr>
<tr>
<td>Knowledge Test</td>
<td>4</td>
<td>Traditional sort of pen and paper test where they were looking at an assessment of knowledge. This measure was only examined in concert with other measures and could not be the sole outcome of a study (otherwise it would be excluded as a non-clinical study).</td>
</tr>
</tbody>
</table>

*Note: “SP” refers to “standardized patient” or “simulated patient”. “CE” refers to “clinical educator”.*
Appendix E

Behavioral Observation Scales

<table>
<thead>
<tr>
<th>Measure</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus-response-consequence completion</td>
<td>Scott et al., (2017)</td>
</tr>
<tr>
<td>Interaction behavior counts</td>
<td>Kaplan &amp; Dreyer (1974)</td>
</tr>
<tr>
<td>Communicative behaviors (specific feedback, one to two-word utterances, comments or statements)</td>
<td>Weltsch &amp; Crowe (2006)</td>
</tr>
<tr>
<td>Peer interaction preschool strategies frequency and variety</td>
<td>Herd (2009)</td>
</tr>
<tr>
<td>Targeted behaviors (clinician explanations, informative feedback, directive responses to neutral/social utterances, pp. 731-32)</td>
<td>Gillam et al. (1990)</td>
</tr>
<tr>
<td>Fidelity Checklist - Targeted PBIS techniques</td>
<td>Lorio et al. (2016)</td>
</tr>
<tr>
<td>Fidelity Checklist - Classroom Pivotal Response Training scale</td>
<td>Donaldson (2015)</td>
</tr>
<tr>
<td>Learning Continuum of Speech &amp; Language Pathologists (LCSLP)</td>
<td>Pamplona (2015)</td>
</tr>
<tr>
<td>Objective Structured Clinical Observation (OSCE)</td>
<td>Zraick et al. (2003)</td>
</tr>
<tr>
<td>Person Centered Behavior Inventory (PCBI)</td>
<td>DiLollo &amp; Favreau (2010)</td>
</tr>
<tr>
<td>Clinical problem solving taxonomy</td>
<td>Moses &amp; Shapiro (1996)</td>
</tr>
<tr>
<td>Situation, Background, Assessment, &amp; Recommendation – Communication Style (SBAR-C)</td>
<td>Towson et al., 2018</td>
</tr>
<tr>
<td>Endoscopic timing (total procedure time &amp; total time in the nose)</td>
<td>Benadom &amp; Potter, 2011</td>
</tr>
<tr>
<td>Type of commitments students agreed to do &amp; completion thereof</td>
<td>Shapiro &amp; Anderson (1988; 1989)</td>
</tr>
</tbody>
</table>

¹The MOSAICS (Weller, 1971) measure was used in other studies that did not meet criteria as well. These studies were largely excluded because they examined differences in talk behaviors between students and clinical educators rather than addressing a specific learning or educational outcome. It was included within this table as the studies referenced used it as an objective measure to correlate with a separate scale.
Appendix F

*Other Measures*

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative analysis with respect to team-building</td>
<td>1</td>
<td>Wainscott (2016)</td>
</tr>
<tr>
<td>Qualitative analysis with respect to team-building</td>
<td>1</td>
<td>Peña &amp; Quinn (2003)</td>
</tr>
<tr>
<td>Number of iPad checkouts following orientation</td>
<td>1</td>
<td>Edwards &amp; Dukhovny (2017)</td>
</tr>
<tr>
<td>Use of a remediation plan or not (note: some were for academic)</td>
<td>1</td>
<td>Troche &amp; Towson (2018)</td>
</tr>
<tr>
<td>Clinical Checkpoint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement with Experts [Simulation Decision]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DecisionSim Time in Seconds to complete [Simulation timing]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation use of strategies as compared to experts</td>
<td>1</td>
<td>Strang &amp; Meyers (1987)</td>
</tr>
<tr>
<td>Talk aloud</td>
<td>1</td>
<td>Boyer (2013)</td>
</tr>
<tr>
<td>Think aloud</td>
<td>1</td>
<td>Ginsberg et al. (2016)</td>
</tr>
</tbody>
</table>