

Consistently inconsistent: An examination of the variability in the identification of emotional disturbance

Kristin Scardamalia¹  | Keisha L. Bentley-Edwards²  | Kairys Grasty³

¹National Center for School Mental Health, School of Medicine, University of Maryland, Baltimore, Maryland

²Duke University, Durham, North Carolina

³Division of Child and Adolescent Psychiatry, School of Medicine, University of Maryland, Baltimore, Maryland

Correspondence

Kristin Scardamalia, National Center for School Mental Health, School of Medicine, University of Maryland, 737 West Lombard St., 4th Floor, Baltimore, MD 21201.
Email: KScardamalia@som.umaryland.edu

Abstract

The federal definition of emotional disturbance (ED) has been heavily criticized as vaguely defined and poorly operationalized yet there has not been a formal analysis of the reliability of the ED criteria. This study examined the reliability of the federal criteria for a special education designation of ED. A total of 179 school psychologists reviewed a mock special education report, made an eligibility determination, and provided information about their eligibility decision. In all, 56 participants found the student met ED eligibility criteria using 16 different combinations of the five criteria. Krippendorff's α , a statistic preferred for content analysis, was calculated as a measure of criterion reliability. Results indicated extremely poor reliability ($\alpha = 0.2011$). These findings demonstrate the importance of redefining the existing criteria.

KEYWORDS

emotional disturbance, reliability, special education assessment

1 | INTRODUCTION

The identification of students with mental health challenges in schools has been problematic and controversial since the establishment of the current special education system. From poorly operationalized eligibility criteria (Donovan & Cross, 2002; Gresham, 2007; Hughes & Bray, 2004) to the disproportionate identification of racial and ethnic minorities in high-incidence categories, especially emotional disturbance (ED; Arnold & Lassmann, 2003; Donovan & Cross, 2002; Patton, 1998; Sullivan, 2017; US GAO, 2013), practitioners and researchers alike have struggled to align the educational definition of ED with the clinical understanding of mental health. Federal education law provides a single definition of ED with five brief criteria; only one of which must be met to determine if a student is emotionally disturbed. Further, identification as ED is associated with poor academic outcomes,

increased risk of contact with the juvenile justice system (Bradley, Doolittle, & Bartolotta, 2008; Cartledge, 1999; Cartledge, Kea, & Simmons-Reed, 2002), and lower rates of employment posthigh school (Bradley et al., 2008). While highlighting the serious concerns with the ED criteria is not new, this paper is the first to present a systematic investigation of examiner decision making from which a reliability estimate of the ED criteria could be derived.

1.1 | Background

In 1975, Public Law 94-142 established special education services and guaranteed a free and appropriate education to all children. Eligibility for special education services requires that a student meet criteria for a qualifying condition in at least one educational disability category which includes learning disability, ED, other health impairments, and intellectual disability. Importantly, the definitions of these school-based disabilities were written by Congress and the criteria for ED have been criticized as poorly operationalized with insufficient detail (Donovan & Cross, 2002; Gresham, 2007; Hughes & Bray, 2004). While changes have been made to other eligibility categories (e.g., learning disability) to address ambiguities (Callinan, Cunningham, & Theiler, 2013), no such revisions have occurred in the definition of ED and no standardized models for determining eligibility have been adopted in federal legislation. This longstanding lack of guidance in the federal definition of ED allows a questionable level of subjectivity to enter into the eligibility determination process which is likely to impact reliability of the criterion themselves. This study provides a systematic examination of the inconsistency in the application of ED criteria including a reliability estimate.

1.2 | ED definition and assessment

Redefinition of diagnostic criteria is hardly unprecedented in the clinical world; in fact, it is far more unusual for a disorder to maintain one set of diagnostic criteria for decades. How is it that against this backdrop of constantly evolving models and definitions in both educational law and clinical diagnostic criteria, there have not been any changes to the educational definition of ED? Per IDEA, to be eligible for special education services as a student with an ED, a student must exhibit one of the five following criteria:

- 1 An inability to learn which cannot be explained by intellectual, sensory, or health factors;
- 2 An inability to build or maintain satisfactory interpersonal relationships with peers and teachers;
- 3 Inappropriate types of behavior or feelings under normal circumstances;
- 4 A general pervasive mood of unhappiness or depression;
- 5 A tendency to develop physical symptoms or fears associated with personal or school problems (IDEA, 2004).

Additional qualifiers must be met indicating that these characteristics have been exhibited over “a long period of time” and to a “marked degree” that “adversely affects” a child’s educational performance, although these phrases have not been defined (Hughes & Bray, 2004). The ED category includes children who have schizophrenia but excludes children who are “socially maladjusted” unless they are found to have “serious emotional disturbance” (IDEA, 2004). The use of social maladjustment (SM) as a disqualifying condition is especially problematic as education law does not provide any definition of SM (Cloth, Evans, Becker, & Paternite, 2014; Hanchon & Allen, 2013). SM has been broadly conceptualized as an externalizing behavior in which the student engages voluntarily and is differentiated from internalizing problems, which are conceptualized as stemming from internal distress over which the student has little or no control (Hinshaw, 1992; Merrell & Walker, 2004; Olympia et al., 2004). However, research has repeatedly found significant overlap between externalizing and internalizing disorders, and assessment procedures do not adequately differentiate between students identified as socially maladjusted and those identified as emotionally disturbed (Cloth et al., 2014; Costenbader & Buntaine, 1999; Donovan & Cross,

2002; McConaughy & Skiba, 1993; Merrell & Walker, 2004). Without clear definitions, the distinction between ED and SM is dependent upon how much self-control the examiner believes the student possesses.

In addition to unclear and absent guidelines, considerable differences exist among the practices of school psychologists when they assess a student to determine ED eligibility (Allen & Hanchon, 2013; Cheramie, Griffin, & Morgan, 2000; Hanchon & Allen, 2013). Data sources for an evaluation can be categorized into five critical elements: classroom observation, teacher interview(s), parent interview(s), student interview, and normative data from rating scales completed by at least two different informants. Research has found that approximately 5% of school psychologists do not consistently include any of the critical data sources listed above and approximately 13% only consistently include one of the five critical data sources. Conversely, approximately 28% consistently include all five sources and almost 30% include four of the five sources (Allen & Hanchon, 2013).

In addition, significant inconsistencies have been found between the types of data that school psychologists rate as the most helpful and the types of data most commonly included in evaluation (Allen & Hanchon, 2013). For example, school psychologists identify interviews with parents as the second most useful source in making an ED determination (Cheramie et al., 2000), and yet, one study found 47.1% of school psychologists included parent interviews in less than half of their evaluations (Allen & Hanchon, 2013). Further, 13 combinations of commonly used assessment measures were investigated for their predictive ability in identifying students with ED. Over the 13 combinations, identification rates ranged from 1.1 to 27.5% (Cluett et al., 1998) indicating that differences in assessment procedures have a notable influence on the identification rate of a disability.

1.3 | Disproportionate representation

The disproportionate representation of racial and ethnic minorities in special education has been a concern since the outset of special education (Arnold & Lassmann, 2003; Donovan & Cross, 2002; Patton, 1998; Sullivan, 2017; US GAO, 2013) despite the growth of diversity-focused training and the adoption of systems designed to improve the quality of referrals to special education (Council for Children with Behavior Disorders, 2013). Three categories of special education have been characterized as high incidence and have been reported to have the highest rates of disproportionality: learning disability, mild intellectual disabilities, and emotional/behavior disorders (Arnold & Lassmann, 2003). Among minority students, African American and American Indian/Alaska Native students have the highest rates of special education placement (US GAO, 2013). African American students in particular have persistently been overrepresented in special education programs for students with learning disabilities, emotional or behavioral disabilities, and mental disabilities (Donovan & Cross, 2002; Patton, 1998; Sullivan, 2017). Their risk is particularly disproportionate for identification as ED (Council for Children with Behavior Disorders, 2013; Donovan & Cross, 2002; Table 1) with twice as many African American students labeled ED when compared with Caucasian counterparts (Aud, Fox, & Kewal-Ramani, 2010; Table 1).

TABLE 1 High incidence eligibility categories by race

Race	Disability category		
	Learning disability (%)	Intellectual disability (%)	Emotional disturbance (%)
Caucasian	3.42	0.60	0.62
African American	5.32	1.56	1.27
Hispanic	4.55	0.59	0.41
Asian/Pacific Islander	1.60	0.39	0.18
American Indian/Alaska Native	7.09	1.01	1.11

Source. Adapted from Aud et al. (2010).

In addition to disproportionate identification of some students, another concern is the altogether missed identification of a group of students who are struggling with significant mental health concerns but whose struggles are not adequately captured under the ED criteria; students who have a history of trauma (Winder, 2015). The types of trauma that a student may experience are highly variable, ranging from repeated exposure to community or domestic violence, physical or sexual abuse, neglect, and exposure to natural disasters. Students' emotional and behavioral manifestations of these experiences are equally variable: difficulty with attention and impulsivity, sleep problems, poor emotional regulation resulting in a tendency to "act out," difficulty with problem solving (Perry, 2009; Van der Kolk, 2005), poor motivation, and difficulty understanding another's perspective (Cole et al., 2005). A cursory glance at this partial list of symptoms makes obvious the difficulty with consistent identification and classification of the effects of trauma under the current disability definitions. The current criteria for ED do not capture the students whose trauma experiences have resulted in hypervigilance, impulsivity, aggression, and acting out when feeling unsafe. Yet, arguably, students who have experienced significant trauma have the greatest need for the types of intervention services associated with ED services in the schools.

1.4 | Consequence of ED identification

Interestingly, despite individualized supports, students served through special education have in the high-incidence categories of mild mental retardation, learning disability, and emotional/behavioral disorders generally have poor outcomes (Arnold & Lassmann, 2003; Bradley et al., 2008; Cartledge, 1999; Cartledge et al., 2002) when compared with nondisabled peers. Students placed in programs for students with ED have the poorest outcomes with lower graduation rates, lower academic achievement, and more encounters with the juvenile justice system (Bradley et al., 2008; Cartledge, 1999; Cartledge et al., 2002). African American males are particularly affected as they are more likely than any other racial or ethnic group to be identified with a disability (Skiba, Poloni-Staudinger, Gallini, Simmons, & Feggins-Azziz, 2006) and in particular with an ED (Cartledge, 1999). Within the category of ED, African American students are overrepresented in more restrictive settings and less likely to be educated in the general education setting than their Caucasian peers (Council for Children with Behavior Disorders, 2013). This pattern has persisted despite the use of individual education plans (IEP) which provide targeted, individualized supports and interventions designed to meet a student's unique needs and allow them to achieve at the same level as their nondisabled peers. One must wonder if the poor outcomes are at least in part due to a failure to correctly identify students whose needs align with the supports historically provided for students identified as ED.

1.5 | Specific criteria

Research on ED identification routinely discusses the eligibility as a single disorder with little to no discussion of the specific criteria used to establish the eligibility. A student must meet only one of the five criteria, which describe different concerns and behaviors (Figure 1). The individual criteria are not operationalized and include

Students must meet one of the following criteria:

1. an inability to learn which cannot be explained by intellectual, sensory, or health factors
2. an inability to build or maintain satisfactory interpersonal relationships with peers and teachers
3. inappropriate types of behavior or feelings under normal circumstances
4. a general pervasive mood of unhappiness or depression
5. a tendency to develop physical symptoms or fears associated with personal or school problems

FIGURE 1 Criteria for emotional disturbance determination (IDEA, 2004)

vague terms such as “normal circumstances,” “inappropriate feelings,” and “satisfactory interpersonal relationships” (Allen & Hanchon, 2013; Donovan & Cross, 2002; Gresham, 2007; Hanchon & Allen, 2013; Hughes & Bray, 2004). Further, the criteria themselves are not well-differentiated from the concept of SM (Cloth et al., 2014; Costenbader & Buntaine, 1999; Donovan & Cross, 2002; McConaughy & Skiba, 1993; Merrell & Walker, 2004). Criteria 2 (inability to build or maintain interpersonal relationships) and Criteria 3 (inappropriate types of behavior or feelings) primarily describe how the student interacts with their environment and the people around them. Criteria 4 (pervasive mood of unhappiness or depression) and Criteria 5 (tendency to develop physical symptoms or fears) are generally interpreted to refer to depressed mood and anxious tendencies, respectively. What then is the distinction between a student who is acting out aggressively due to “inappropriate types of behaviors” and the student who is socially maladjusted? The significant overlap between ED and the disqualifying label of SM, in combination with the lack of clear definition of the individual criteria, has potentially serious implications for the reliability of the ED criteria. The shared characteristics of ED and SM muddies the distinction between these two theoretical constructs and undermines the reliability of the definition by increasing opportunity for clinician inconsistency in identification of ED (i.e., error).

An examination of the individual criteria is appropriate for several reasons. First and foremost, the criteria themselves *are the definition* of ED. The federal definition of ED does not contain an overarching definition, and can only be understood as a function of the individual criteria. Consider the areas under learning disabilities; knowing that a student has a learning disability tells us very little about the needs of that student. However, knowing that a student has a learning disability in reading comprehension provides a starting point for conceptualizing the student’s needs. While identifying the category of reading comprehension itself is not sufficient to design individualized instructional supports, it is a necessary starting point.

The specific ED criteria, or combination of criteria, represent an important piece of the case conceptualization and in effect, labels the relevant behaviors that the school psychologist identified as contributing to a disabling condition. Labels are important because they provide a context through which we interpret the actions of others (Noblit, Paul, & Schlechty, 1991). Further, the criteria describe a wide range of emotional and behavioral concerns in school and interventions need to align with the student’s needs to be effective (Benner, Beaudoin, Chen, Davis, & Ralston, 2010). The initial reason for identification of an ED disability frames the core of the concerns, providing the context through which the student’s behavior is interpreted and thus provides a starting point for conceptualizing a student’s needs and designing interventions; clearly interventions designed to alleviate physical symptoms associated with school would have little effectiveness in addressing an inability to maintain relationships. While the final eligibility decision is made by a multidisciplinary team, school psychologists are responsible for the formal evaluation of emotional and psychological processes including protocol selection, informant interviews, and a recommendation for or against eligibility. This central role in the evaluation process gives the school psychologist substantial influence on decisions made by the team (Hanchon & Allen, 2013).

Identification of core issues leading to an ED eligibility becomes especially crucial in the case of manifestation determinations. Manifestation determination meetings are conducted when a disciplinary action would lead to a change in placement for a special education student. Punitive changes in placement typically involve a suspension or placement in an alternative education program. Special education law requires that IEP teams consider whether the student’s disability had a direct influence or caused the behavior for which the student is being disciplined (IDEA, 2004). If the team determines that the disability did have a direct influence, then the student’s educational placement may not be changed. If the behavior in question was not linked to the disability, then the student may receive the same consequences as a general education student, including suspension. In these circumstances, a valid and reliable identification of the type of emotional and behavioral concerns exhibited by a student can have a large impact on the trajectory of that student’s school career.

A close examination of the performance of the ED criteria is notably lacking from the literature—whether a consideration of the reliability and validity of the eligibility criteria or as guides for intervention strategies. Given the longstanding concerns with the definition of ED and the ambiguity of the individual criteria, a formal analysis of

the reliability of the criteria is warranted. This paper presents findings from an experimental study of decision making by school psychologists that examined factors which influenced the special education eligibility decisions made by evaluators.

1.6 | Current study

The current study is post hoc analysis derived from a larger study which examined the influence of implicit social cognitions on eligibility decisions made by school psychologists. To examine influences on decision making, currently practicing school psychologists in Texas were recruited to participate in an online study for which they reviewed a mock special education evaluation report, determined if the student met criteria for an educational disability, and completed a questionnaire. A review of the results suggested a concerning amount of inconsistency in the application of criteria for ED. The hypotheses of the larger study were focused on the *magnitude* of the perceived concern and whether it rose to the level of an educational disability; however, initial reviews of the data suggested unanticipated variability in the criteria evaluators used to establish the eligibility for ED as well as conflicting perceptions of the student's functioning, prompting a post hoc reliability analysis. While the mock evaluation report used for the study intentionally presented a "borderline" case, (one in which an evaluator could decide yes-ED or no-ED), it was anticipated that of those psychologists who decided the student met criteria, most would select similar criteria to determine eligibility. This paper presents the results of a reliability estimate of the ED criteria.

2 | METHODS

All study materials were presented via a secure website. Participants were presented with one of six randomly assigned mock full and individual initial evaluations (FIIE). Participants reviewed the report and indicated their eligibility determination by selecting one or more ED criteria or by indicating the student did not meet criteria. Participants then completed a brief questionnaire which included demographic information, questions about the sufficiency of the information in the report, and an open-ended question about why the participant felt the student did or did not meet criteria for ED.

TABLE 2 Study participant characteristics

	<i>n</i>	Percentage of sample	Mean	<i>SD</i>	Range
Age	167	-	41.30	11.98	20-75
Years of experience	174	-	10.99	7.98	1-38
Race	171	-	-	-	-
Caucasian	129	73.7	-	-	-
African American	11	6.3	-	-	-
Hispanic/Latino(a)	20	11.4	-	-	-
Asian	1	0.6	-	-	-
Bi/multiracial	6	3.4	-	-	-
Other	4	2.3	-	-	-
Degree level	174	--	-	-	-
Masters	65	37.1	-	-	-
Specialist	78	44.6	-	-	-
Psy.D.	1	0.6	-	-	-
Ph.D.	30	17.2	-	-	-

2.1 | Participants

One hundred and seventy-nine licensed specialists in school psychology (LSSP) were recruited between January and June 2016. A list of 3,240 LSSPs with current licenses was obtained from the state licensing board. An email address was available for approximately 75% of the LSSPs; a mailing address was available for the remainder. LSSPs were contacted via email on three occasions; a postcard invitation was mailed to the remaining LSSPs on two occasions. Four respondents did not provide an eligibility determination and were excluded from the analyses. Participant characteristics (Table 2) closely matched the results of the 2015 National Association of School Psychologists (NASP) demographic survey of school psychologists (Walcott & Hyson, 2018) in terms of age and years of experience. Study participants had a mean age of 41.3 compared with the NASP reported mean age of 42.4. Study participants had a mean years of experience of 10.99 compared with the NASP reported 12.2 mean years of experience. Participant characteristics were similar but less closely matched for race and degree level. The study sample included fewer Caucasian participants (73.7%) than the NASP sample (88.2%) and more participants identifying as Hispanic or Latino(a) (11.4%) than the NASP sample (6%). Study participants were much more likely to have a master's degree (37.1%) than the NASP sample (19%) and less likely to have a doctoral degree (17.8%) than the NASP sample (25.2%).

2.2 | Instruments

2.2.1 | Mock FIIE

A mock FIIE was written by the researcher in conjunction with a focus group of four LSSPs; all practicing in public schools and all of whom had a minimum 10 years' experience in the evaluation of ED. Each LSSP estimated they complete between 15 and 20 evaluations of ED per academic year. The mock FIIE underwent multiple revisions but was not piloted before use in the current study. To meet the goals of the larger study, one mock FIIE was developed and varied only on race (three conditions) and socioeconomic status (two conditions) to create six versions. One of three race conditions was described in the report header and in the sociological section. One of two socioeconomic status conditions was described in the sociological section. The analyses for this study used reports from any of the six versions in which the participant determined the student met ED criteria; identification rates for ED were not significantly different between the six versions.

The mock FIIE described a male student who was struggling academically and behaviorally. The report was presented in the same format as a typical FIIE in Texas, including educational history, language proficiency, physical/health information, and sociological information. Assessment information included a review of educational records, parent, and teacher interviews, a student interview, a Behavior Assessment System for Children, second edition (Reynolds & Kamphaus, 2004) completed by a parent, teacher, and a mentor, along with the following self-report measures: Piers-Harris two Children Self Concept Scale (Piers & Herzberg, 2002), Children's Depression Inventory (Kovacs, 1985), and the Adolescent Anger Rating Scale (Burney, 2001).

2.2.2 | Emotional disturbance determination

After reviewing the report, evaluators were asked to make a determination of eligibility by selecting one or more of the eligibility criteria or choosing a "does not qualify" option (Figure 1). This "yes or no" decision is analogous to decision making as it happens in schools as evaluators must ultimately choose to qualify a student or to not qualify.

2.2.3 | FIIE and demographics questionnaire

A researcher-developed questionnaire provided participants the opportunity to voice concerns about the report, explain their decision, and provide demographic information. Participants were presented with the

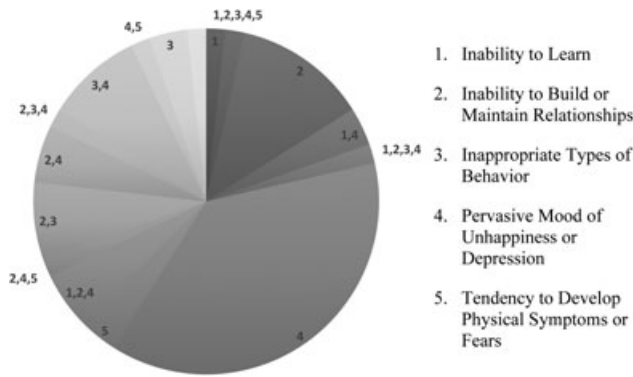


FIGURE 2 Frequency of ED criteria combinations. This figure shows the different combinations used by participants to establish an ED eligibility. ED: emotional disturbance

questionnaire immediately after making an eligibility determination. The first question asked participants to indicate if they felt they had sufficient information to make an eligibility decision by selecting yes or no. Participants were then given the option to write in any additional information they wished had been available. Next, participants responded to an open-ended question asking them to briefly describe in their own words why the student did or did not meet eligibility criteria. Finally, participants were asked to provide demographic information including race, age, level of training, years of LSSP experience, average number of ED evaluations completed per school year, and the location of their current school district (defined as rural, suburban, or urban).

2.3 | Analytic strategy

First, an overall percentage of respondents who felt the student met criteria for ED was calculated. Next, the specific criterion, or combinations of criteria, used to make the ED determination were reviewed for consistency. The total number of times each unique combination was used was tallied and examined for patterns (Figure 2). The surprisingly high number of different combinations of criteria was followed by a review of the participants' description of why they felt the student did or did not meet ED criteria. A number of contradictory statements were found in the review (Figure 3), further supporting concerns for the level of variability in responses. A post hoc analysis of the reliability of the ED criteria was then conducted.

Participant determined student <i>did meet</i> ED Criteria	Participant determined student <i>did not meet</i> ED Criteria
"he does not have many friends"	"capable of making friends"
"consistently demonstrated behavioral concerns for an extended period of time"	"behavior and academic difficulties were not significant"
"significant difficulties getting along with peers"	"appropriate peer relationships"
"consistency in reported issues across settings"	"there is little consistency between raters"
"significant moodiness, episodic anger"	"absence of significant depression, anxiety"

FIGURE 3 Participant written statements. This figure demonstrates the range of statements written by participants in response to open-ended writing prompt. ED: emotional disturbance

To determine the reliability of the criteria themselves, Krippendorff's α ($K\alpha$), a statistic preferred for content analysis, was used as a measure of criterion reliability. $K\alpha$ is an appropriate reliability estimate for decisions made at any level of measurement, including nominal data, and any number of observers (Hayes & Krippendorff, 2007). Further, $K\alpha$ is a measure of the reliability of the variable, not respondents, and therefore provides information as to the robustness of the ED criteria (Hayes & Krippendorff, 2007). Only responses that indicated the student met criteria for ED were used for the $K\alpha$ analysis. For each respondent, all five eligibility criteria were coded separately; 1 if the criteria was selected and 0 if the criteria was not selected. Due to the relatively small sample size ($n = 56$), bootstrapping was utilized to provide a more robust measure. Bootstrapping is a resampling technique used to create a larger sample while maintaining the original sample's proportions which provides a more accurate estimation of the indirect effects tested (Agresti, 2002; Preacher, Rucker, & Hayes, 2007). Bootstrapping produces a confidence interval based on the 95th percentile and is considered valid if the upper and lower bounds do not include zero; that is if both the upper and lower bounds are positive or negative. If zero is captured in the confidence interval, the model is considered invalid (Preacher et al., 2007). One thousand resamples were produced for the $K\alpha$ analysis.

3 | RESULTS

Data were gathered from 179 practicing LSSPs. Four respondents did not provide an eligibility determination and were excluded from the analyses. Of the 175 eligibility determinations, 56 participants indicated that the student met criteria for ED and 119 indicated the student did not meet criteria.

3.1 | Inconsistency in application of criteria

Results of this study suggest that the 56 participants who determined the student met the eligibility for ED had vastly different interpretations of assessment results and the appropriate application of the five criteria for ED. All five criteria were used at least once as a sole determinant and an additional 11 unique combinations of criteria were used to establish eligibility (Figure 2). Criteria 4 (pervasive sense of unhappiness or depression) was used most frequently and appeared in 39 of the 56 positive determinations of ED, both as a sole determinant and in combination with other criteria. The degree of variation within the group of participants who determined the student met ED criteria indicates that even when participants agreed the student was ED, they disagreed on what characteristics the student was exhibiting that indicated ED.

3.2 | Inconsistency in qualitative descriptions

The study questionnaire included a prompt for participants to briefly describe in their own words why they felt the student did or did not meet ED criteria. Perhaps not surprisingly, participants who felt the student met criteria described the student differently than the participants who felt the student did not meet criteria. However, if all participants were interpreting the evaluation data in relatively the same manner, then the difference in descriptions should primarily be based on the magnitude of the perceived issue. A review of the written statements revealed a wide range of perceptions about the student even though the evaluation information, including teacher reports, assessment data, educational history, and interview information, was the same. In other words, although respondents looked at the same data, some respondents determined the student was depressed, some thought he was anxious, while others felt inappropriate behavior was the primary concern. This inconsistency becomes even greater when you consider that twice as many respondents felt the student did not meet criteria, often arguing the exact opposite as reported above; that the report failed to establish the presence of depression or anxiety, that the student appeared to get along with others, and even that a lack of academic progress was not established. To

highlight the stark differences in the perception of this student, a sample of statements written by participants (direct quotes) are presented (Figure 3).

3.3 | Reliability estimate

Given the unanticipated number of different combinations of criteria and contradictory statements made by participants, a measure of criterion reliability was calculated. Results of $K\alpha$ reliability estimate indicated poor reliability ($\alpha = 0.2011$) and a confidence interval was produced that included zero indicating an invalid model (see Table 3). $K\alpha$ also produced a probability of failure to achieve a minimum α ; in other words, the percent chance that the reliability estimate would fail to meet a target α . Standard calculations produce a percent chance of failure for target α 's of 0.90, 0.80, 0.70, 0.60 and 0.50. Results indicated a 100% chance of failure to achieve a minimum α of 0.6 or greater, and 99.7% chance of failure to produce a minimum α of 0.50. These results indicate a zero probability that the criteria could have met a target α of at least 0.6 and a very high probably of failure to reach even an α of 0.50. Taken together, results indicated that the current criteria are not a reliable measure of ED.

4 | DISCUSSION

Results of this study support the criticism of the ambiguous and poorly operationalized definition of ED. Most literature has focused on the significant overlap between ED and SM or the lack of clarity of phrases such as "marked degree" or "appropriate behavior" in the definition of ED (Hughes & Bray, 2004). However, at a more foundational level, examiners appear to have different interpretations of what defines ED, and more surprisingly, different interpretations of the same assessment results. The 56 respondents who felt the student met eligibility used all five criteria individually and an additional 11 different combinations of criteria to establish eligibility. A review of statements made by study participants revealed different and sometimes opposing perceptions of the student and potential emotional or behavioral concerns. This inconsistency was quantified by computing a reliability estimate of the ED criteria ($\alpha = 0.2011$) which indicated a concerning lack of reliability in the criteria themselves.

First and foremost, results of this study support the calls by researchers to re-evaluate and update the definition of ED and to standardize the evaluation process used to make ED determinations (Allen & Hanchon, 2013; Donovan & Cross, 2002; Gresham, 2007; Hanchon & Allen, 2013; Hughes & Bray, 2004; Merrell & Walker, 2004). There is precedent within the educational system as methods for determining learning disabilities have

TABLE 3 Results of Krippendorff's α for ED criteria

Criterion reliability estimate					
α	LL 95%CI	UL 95%CI	Units	Observers	Pairs
0.2011	-0.0110	0.4256	5.00	56.00	7700.00
Probability (q) of failure to achieve a target α					
Target α	q				
0.9000	1.000				
0.8000	1.000				
0.7000	1.000				
0.6000	1.000				
0.5000	0.997				

Note. CI: confidence interval; ED: emotional disturbance; LL: -0.0110; UL: 0.4256.

changed substantially due to insufficient operationalization (Callinan et al., 2013) and precedent in the clinical world for revision of clinical diagnostic categories and criteria to align with the most recent research. Yet, the original definition of ED from 1975 has been maintained.

The poor reliability of the ED criteria is especially problematic when viewed in the context of the overidentification of some groups, the under identification of others, and the failure to address the symptoms of exposure to trauma that many students exhibit in our schools. Intervention strategies are only effective when they are aligned with the student's needs (Benner et al., 2010) and accurately identifying which behavior (s) the student is demonstrating that meet criteria for ED is a crucial first step to help the IEP team select appropriate strategies. Among students served through special education, students who are identified as ED have the poorest outcomes with lower graduation rates, lower academic achievement, and more encounters with the juvenile justice system (Bradley et al., 2008; Cartledge, 1999; Cartledge et al., 2002). Is it possible that the inconsistencies in interpretation of assessment data and the unreliability of the ED criteria are a major contributor to poor outcomes through misalignment of needs and interventions? Research on the ED criteria themselves is lacking in the literature but may provide some guidance on improving the outcomes for students identified as ED. In addition, research on how LSSPs conceptualize ED will help illuminate the current trends in working definitions and provide guidance on how to better operationalize the definition, thus improving the ability to reliably assess for the presence of ED.

4.1 | Limitations and future directions

The generalizability of this study is limited due to the small sample size. However, the use of bootstrapping in the statistical analysis offers a clear picture of the inadequacies of the established criterion for ED. Additionally, all of the participants were practicing in the same state (Texas). Study participant demographics closely match a national sample in terms of age and years of experience but are less closely matched on race and degree level. Study participants were more likely to hold a master's degree versus a doctoral degree which may indicate a difference in initial training.

Additional limitations exist due to the study methodology. While the mock FIIE was developed with a team of experienced school psychologists, the mock report was not pilot tested. Future iterations of this study would be strengthened by a rigorous pilot test of the stimulus. In addition, school psychologists read the mock FIIE online and while they reviewed assessment results, direct observation of the student was not available. Classroom observations are an important contextual factor in interpreting student behavior and the effect on eligibility determinations is unclear. Similarly, school psychologists were not in the testing sessions with the student and did not have the opportunity to observe the student involved in assessments tasks. Last, final determinations of eligibility are made by the IEP team, not the school psychologist alone.

One manner of improving consistency in eligibility determinations is to promote the codification of standardized assessment protocols with clear guidelines for meeting criteria. While NASP has published best practice guidelines, school psychologists are not legally obligated to follow those guidelines. Current research reveals that in practice there is little consistency in the types of information gathered and the measures used by school psychologists (Allen & Hanchon, 2013; Cheramie et al., 2000; Hanchon & Allen, 2013). Research has also demonstrated that the types of data used in considering an eligibility have a measurable impact on the final determination (Cluett et al., 1998). Increasing consistency in assessment procedures will help align decision making by school psychologists.

And finally, this paper adds to the many calls for the revision of the ED criteria at the Federal level. While beyond the scope of this paper, several alternative definitions and processes have been proposed. Interested readers are directed to Merrell and Walker's (2004) discussion of the alternate definition proposed by the National Mental Health and Special Education Coalition and of a proposed reclassification of EBD. Additionally, an alternate process based on Response to Intervention methodologies has been proposed by Maag and Katsiyannis (2008).

5 | CONCLUSION

Results of this study indicate a problematic level of inconsistency in the interpretation and application of the criteria for ED. Given the poor outcomes associated with the identification of ED, it is concerning that the foundation for decision making, the criteria themselves, have been found to perform poorly in criterion reliability testing. While the troubling ambiguity of the ED definition has been long noted in the literature, this study provides a quantification of the poor reliability and supports the many calls by school psychologists to redefine this category of special education eligibility.

ORCID

Kristin Scardamalia  <http://orcid.org/0000-0001-7903-0533>

Keisha L. Bentley-Edwards  <http://orcid.org/0000-0001-8430-4850>

REFERENCES

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.
- Allen, R. A., & Hanchon, T. A. (2013). What can we learn from school-based emotional disturbance assessment practices? Implications for practice and preparation in school psychology. *Psychology in the Schools, 50*, 290–299. <https://doi.org/10.1002/pits.21671>
- Arnold, M., & Lassmann, M. (2003). Overrepresentation of minority students in special education. *Education, 124*, 230–236. Retrieved from . <https://www.projectinnovation.com/education.html>
- Aud, S., Fox, F., Kewal-Ramani, A. (2010). *Status and trends in the education of racial and ethnic groups*, US Department of Education, National Center for Education Statistics (NCES 2010-015). Washington, DC: US Government Printing Office. Retrieved from <http://nces.ed.gov/>
- Benner, G. J., Beaudoin, K. M., Chen, P. Y., Davis, C., & Ralston, N. C. (2010). The impact of intensive positive behavioral supports on the behavioral functioning of students with emotional disturbance: How much does fidelity matter? *Journal of Behavior Assessment and Intervention in Children, 1*, 85–100. <https://doi.org/10.1037/h0100361>
- Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education, 17*, 4–23. <https://doi.org/10.1007/s10864-007-9058-6>
- Burney, D. (2001). *Adolescent Anger Rating Scale professional manual*. Odessa, FL: Psychological Assessment Resources.
- Callinan, S., Cunningham, E., & Theiler, S. (2013). Revisiting discrepancy theory in learning disabilities: What went wrong and why we should go back. *Australian Journal of Guidance and Counselling, 23*, 1–17. <https://doi.org/10.1017/jgc.2012.22>
- Cartledge, G. (1999). African-American males and serious emotional disturbance: Some personal perspectives. *Behavioral Disorders, 25*, 76–80. <https://doi.org/10.1177/019874299902500101>
- Cartledge, G., Kea, C., & Simmons-Reed, E. (2002). Serving culturally diverse children with serious emotional disturbance and their families. *Journal of Child and Family Studies, 11*, 113–126. <https://doi.org/10.1023/A:1014775813711>
- Cherame, G. M., Griffin, K. M., & Morgan, T. (2000). Usefulness of assessment techniques in assessing classification for emotional disturbance and generating classroom recommendations. *Perceptual and Motor Skills, 90*, 250–252. <https://doi.org/10.2466/pms.2000.90.1.250>
- Cloth, A. H., Evans, S. W., Becker, S. P., & Paternite, C. E. (2014). Social maladjustment and special education: State regulations and continued controversy. *Journal of Emotional and Behavioral Disorders, 22*, 214–224. <https://doi.org/10.1177/1063426613487405>
- Cluett, S. E., Forness, S. R., Ramey, C. T., Hsu, C., Kavale, K. A., & Gresham, F. M. (1998). Consequences of differential diagnostic criteria on identification rates of children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders, 6*, 130–140. <https://doi.org/10.1177/106342669800600301>
- Cole, S., Greenwald O'Brien, J., Geron Gadd, M., Ristuccia, J., Luray Wallace, D., & Gregory, M. (2005). *Helping traumatized children learn: Supportive school environments for children traumatized by family violence*. Boston, MA: Massachusetts Advocates for Children, Trauma and Learning Policy Initiative.
- Costenbader, V., & Buntaine, R. (1999). Diagnostic discrimination between social maladjustment and emotional disturbance: An empirical study. *Journal of Emotional and Behavioral Disorders, 7*, 2–10. <https://doi.org/10.1177/106342669900700101>
- Council for Children with Behavior Disorders (2013). CCBBD's position summary on federal policy on disproportionality in special education. *Behavioral Disorders, 38*, 108–120. <https://doi.org/10.1177/019874291303800202>

- Donovan, M., & Cross, C. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- Gresham, F. M. (2007). Response to intervention and emotional and behavioral disorders: Best practices in assessment for intervention. *Assessment for Effective Intervention*, 32, 214–222. <https://doi.org/10.1177/15345084070320040301>
- Hanchon, T. A., & Allen, R. A. (2013). Identifying students with emotional disturbance: School psychologists' practices and perceptions. *Psychology in the Schools*, 50, 193–208. <https://doi.org/10.1002/pits.21668>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1, 77–89. <https://doi.org/10.1080/19312450709336664>
- Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, 111, 127–155. <https://doi.org/10.1037/0033-2909.111.1.127>
- Hughes, T. L., & Bray, M. A. (2004). Differentiation of emotional disturbance and social maladjustment: Introduction to the special issue. *Psychology in the Schools*, 41, 819–821. <https://doi.org/10.1002/pits.20038>
- IDEA, 2004. Individuals with Disabilities Education Act of 1990, Part B §300.8
- Van der Kolk, B. A. (2005). Developmental trauma disorder: Toward a rational diagnosis for children with complex trauma histories. *Psychiatric Annals*, 35, 401–408. <https://doi.org/10.3928/00485713-20050501-06>
- Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin*, 21, 995–998.
- Maag, J. W., & Katsiyannis, A. (2008). The medical model to block eligibility for students with EBD: A response to intervention alternative. *Behavioral Disorders*, 33, 184–194. <https://doi.org/10.1177/019874290803300305>
- McConaughy, S., & Skiba, R. (1993). Comorbidity of externalizing and internalizing disorders. *School Psychology Review*, 22, 421–436. <http://psycnet.apa.org/record/1994-17824-001>
- Merrell, K. W., & Walker, H. M. (2004). Deconstructing a definition: Social maladjustment versus emotional disturbance and moving the EBD field forward. *Psychology in the Schools*, 41, 899–910. <https://doi.org/10.1002/pits.20046>
- Noblit, G. W., Paul, J. L., & Schlechty, P. (1991). The social and cultural construction of emotional disturbance. In Paul, J. L., & Epanchin, B. C. (Eds.), *Educating emotionally disturbed children and youth: Theories and practices for teachers* (2nd ed., pp. 218–242). Columbus, OH: Merrill.
- Olympia, D., Farley, M., Christiansen, E., Pettersson, H., Jenson, W., & Clark, E. (2004). Social maladjustment and students with behavioral and emotional disorders: Revisiting basic assumptions and assessment issues. *Psychology in the Schools*, 41, 835–847. <https://doi.org/10.1002/pits.20040>
- Patton, J. M. (1998). The disproportionate representation of African Americans in special education: Looking behind the curtains for understanding and solutions. *The Journal of Special Education*, 32, 25–31. <https://doi.org/10.1177/002246699803200104>
- Perry, B. D. (2009). Examining child maltreatment through a neurodevelopmental lens: Clinical applications of the neurosequential model of therapeutics. *Journal of Loss and Trauma*, 14, 240–255. <https://doi.org/10.1080/15325020903004350>
- Piers, E., & Herzberg, D. (2002). *Piers-Harris Children's Self-Concept Scale—Second edition manual*. Los Angeles, CA: Western Psychological Services.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypothesis: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42, 185–227. <https://doi.org/10.1080/00273170701341316>
- Reynolds, C., & Kamphaus, R. (2004). *BASC-2: Behavior assessment system for children* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Skiba, R. J., Poloni-Staudinger, L., Gallini, S., Simmons, A. B., & Feggins-Azziz, R. (2006). Disparate access: The disproportionality of African American students with disabilities across educational environments. *Exceptional Children*, 72, 411–424. <https://doi.org/10.1177/001440290607200402>
- Sullivan, A. L. (2017). Wading through quicksand: Making sense of minority disproportionality in identification of emotional disturbance. *Behavioral Disorders*, 43, 244–252. <https://doi.org/10.1177/0198742917732360>
- US Government Accountability Office Report to the Chairman, Committee on Health, Education, Labor, and Pensions, U.S. Senate. (February 2013). *Individuals with disabilities education act: Standards needed to improve identification of racial and ethnic overrepresentation in special education* (GAO-13-137). <http://www.gao.gov/products/GAO-13-137>
- Walcott, C. M., & Hyson, D. (2018). *Results from the NASP 2015 membership survey, part one: Demographics and employment conditions [Research report]*. Bethesda, MD: National Association of School Psychologists.
- Winder, F. (2015). Childhood trauma and special education: Why the "IDEA" is failing today's impacted youth. *Hofstra law review*, 44, 601–634. Retrieved from <http://scholarlycommons.law.hofstra.edu/hlr/vol44/iss2/18>

How to cite this article: Scardamalia K, Bentley-Edwards KL, Grasty K. Consistently inconsistent: An examination of the variability in the identification of emotional disturbance. *Psychol Schs*. 2019;56:569–581. <https://doi.org/10.1002/pits.22213>