

Exploring the Ability of Educators to Identify Behaviors Indicative of Emerging Psychopathologies in Elementary School Students: Assessing the Use of a Novel Vignette Measure

Canadian Journal of School Psychology
2024, Vol. 39(1) 72–86
© The Authors 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08295735241228069
journals.sagepub.com/home/cjs



Damian Page¹  and Todd Cunningham¹

Abstract

The present study sought to assess the ability of teachers to identify emerging mental health disorders through a novel vignette measure. Canadian certified primary grade teachers ($N = 101$) completed a survey that included a novel vignette measure. Participants rated the severity of fictitious student behaviors depicted in several vignettes and their accuracy was calculated based on how closely their ratings matched the severity of symptoms depicted. Accuracy estimates derived through this measure differed considerably from previous vignette measure paradigms, producing much lower estimates of identification accuracy. A binomial logistics regression indicated that neither the gender nor pathology depicted in the vignettes significantly influenced rating accuracy. This novel vignette measure may represent a quick and effective means of assessing the accuracy of teachers in identifying emerging mental health disorders in their students.

Keywords

mental health literacy, mental health, elementary students, psychopathology, teachers, participants

¹University of Toronto, Toronto, ON, Canada

Corresponding Author:

Damian Page, Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, St George Campus, 200 Balliol Street, Apt 902, Toronto, ON M4S 1C6, Canada.

Email: damian.page@mail.utoronto.ca

Background

Identifying emerging mental disorders in children is a complex skill that requires differentiating between behaviors within a typical development range and those indicative of emerging psychopathologies. These discriminatory judgments are difficult given the possibility of non-typical manifestations of mental disorders in childhood, such as depression, which can present in children as physical symptoms without the stereotypical psychological effects (Stewart, 2003). The process by which teachers identify students in their class as experiencing the onset of a mental disorders represents the first step in the help-seeking pathway for students.

Research assessing the accuracy by which teachers recognize the incidence of mental disorders in their students has primarily relied on two methods, classroom nominations of students suspected of experiencing mental health distress, and vignette measure designs (Cunningham & Suldo, 2014; M. Green et al., 1996; J. G. Green et al., 2018; Loades & Mastroyannopoulou, 2010; Splett et al., 2019). In the former method, a sample of teachers are asked to nominate students in their class that they suspect are presenting symptoms indicative of a particular psychopathology. These nominations are then compared to clinical symptomatic elevations in students, obtained using standardized clinical measurement scales. Studies examining the accuracy of teacher nominations in identifying students with elevations in depression and anxiety symptoms have found miss rates as high as 50 and 60% respectively (Cunningham & Suldo, 2014). Although ecologically valid, classroom nomination methods are complex and costly, thereby limiting the sample size of these studies. In comparison, vignette measures are easy to administer and require less complex research designs.

Vignette measure designs involve presenting participants with a short-written description of a fictional person exhibiting behaviors indicative of a mental disorder and then asking them to provide their impressions of the behaviors depicted (e.g., level of concern, diagnostic label, recommended referral actions, beliefs about treatments, outcomes, causes, risk factors, the health status of the fictitious individual, and stigmatizing attitudes toward the person depicted in the vignette). In comparison to relatively low estimates of identification accuracy derived from classroom nomination methods, vignette measure designs can produce accuracy estimates as high as 90 to 100% (Splett et al., 2019). This incongruence between classroom nomination and vignette designs to assess the identification accuracy of teachers, is likely a reflection of the methodological constraints inherent to these vignette measure designs.

The vignette methodology employed by Jorm et al. (1997) and others (J. Green et al., 2018; M. Green et al., 1996; Loades & Mastroyannopoulou, 2010) reflects the categorical nosology upon which the Diagnostic and Statistics Manual 5 (DSM-5) is built, and through which mental health disorders are diagnosed in Canada (American Psychiatric Association, 2013). Within this theoretical framework, criteria for the diagnosis of a mental health disorders are either met or rejected based on the frequency, intensity, onset, and duration of specific symptoms. Vignette measures that utilize a problem present/absent dichotomy require a less nuanced consideration of the

developmental range of typical student behaviors and the extraneous factors that may drive such behavioral presentations. Instead, participants are asked to simply indicate if the person depicted does or does not have a mental health disorder, problem, or concern. Ignoring the dimensional severity through which most psychopathologies present, may therefore artificially inflate identification accuracy estimates derived through vignette measures with dichotomized response options. This may explain why identification accuracy estimates vary so significantly between vignette and classroom nomination measure designs.

A study by Splett et al. (2019) examined the ability of teachers to accurately identify mental health concerns within vignettes depicting elementary school-age children presenting with moderate or severe externalizing or internalizing symptoms. Teachers were able to accurately identify the presence of problematic internalizing and externalizing behaviors when they were severe but struggled to accurately identify moderate and subclinical levels of dysfunction (Splett et al., 2019). These results are concerning because they suggest that educators may fail to identify the emergence of mental health problems at the subclinical level, which represents an obstacle to early detection and intervention.

Despite the methodological shortcomings of vignette measures in approximating the real world decisions educators make in differentiating typical from atypical student behaviors, they do provide more experimental control, allowing researchers to systematically modify student variables within the vignettes (i.e., gender and pathology expression type). Studies using vignette measures to assess teacher perceptions indicate that the gender and pathology presentation (e.g., externalizing, or internalizing disorders) of the depicted student can significantly influence levels of concern or referral decisions (M. Green et al., 1996; J. Green et al., 2018; Loades & Mastroyannopoulou, 2010; Splett et al., 2019). A study by Loades and Mastroyannopoulou found that teachers were more accurate in their identification of mental health problems when rating vignettes describing boys with externalizing behaviors and girls with internalizing behaviors than vice versa. These results suggest the presence of a gender bias in teacher appraisals of behavioral severity, which may reflect their real world teaching experiences. Emotional disorders are more common in girls, while behavioral disorders are more prevalent in boys (Maughan et al, 2004; Santa Lucia et al., 2000). The presence of response biases driven by the gender and pathology of depicted students within vignette measures, which aligns with real world prevalence rates of externalizing and internalizing symptoms in boys and girls, seems to support the ecological validity of vignette measures.

The Present Study

The current study attempted to expand upon the vignette measure design popularized by Jorm et al. (1997), by incorporating a dimensional severity structure within the vignettes. Accurately assessing the ability of teachers to identify emerging mental health disorders in students and refer them to mental health services requires differentiating between behaviors within and beyond a typical developmental range. This

process of discrimination has until now been limited by the dichotomized response modality of previous vignette measures, in which a mental health disorder is either present or absent without considerations to subclinical levels of dysfunction and impairment. The construction of a novel vignette measure for this study attempted to address the shortcomings of previous diagnostic vignette measures by incorporating elements of dimensional models of psychopathology that better capture subclinical levels of problem behaviors.

To this end, two research questions were explored. First, would the gender and pathology biases observed in previous studies employing a vignette measure design also be observed in participants' responses to our novel vignette measure. Previous research utilizing vignette methods has shown that teachers' ratings of concern or identification of problem behavior is significantly influenced by the gender and type of pathology depicted in the vignette. (M. Green et al., 1996; J. Green et al., 2018; Loades & Mastroyannopoulou, 2010). The researchers hypothesized that participants' discriminatory accuracy would be significantly influenced by the depicted gender and pathology expressions within the vignettes.

Secondly, would participants' discrimination accuracy be similar to accuracy estimates derived through previous vignette measure studies. Incorporating a dimensional severity structure as well as thematic components was done to approximate the kind of information teachers have access to when making identifications or referral decisions. Therefore, the researchers hypothesized that participant's identification accuracy would be more similar to accuracy estimates derived through non-vignette measures, such as classroom nomination measure designs (Cunningham & Suldo, 2014).

Methodology

Design

To address the aforementioned research questions, a quantitative cross-sectional research design was employed. The participants, Canadian certified elementary school teachers of Grades 1 to 6, completed an online survey consisting of a novel vignette measure, a 35-item mental health literacy scale (MHLS) developed by O'Connor and Casey (2015) and several demographic questions. For the purpose of this study and in answering the aforementioned research questions, only data relating to participants performance on the vignette measure is discussed.

This study design utilizing online data collection was selected in order to obtain a large, diversified data set at low cost. Ethical approval for this study was obtained through the University of Toronto Human Research Ethics Program review board.

Participants

Participants were certified Canadian elementary teachers currently teaching elementary school Grades 1 to 6. Participants were recruited through targeted email invitations to professional teaching networks in collaboration with the Ontario Institute

for Studies in Educations' Continuing Professional Learning department. Further participant recruitment was conducted through a professional survey dissemination service provided by Qualtrics.

Measures

The creation of a novel vignette measure for this study was done through an iterative process, incorporating feedback from faculty and graduate students in the School and Clinical Child Psychology program at OISE, certified Ontario teachers, and registered psychologists. Data was collected and analyzed to explore the influence of gender, pathology, and symptom severity on reviewers' responses. This pilot data was then used to inform subsequent iterations of the vignettes. Depictions of behaviors presented in the vignettes were designed to accurately represent the kinds of student behaviors commonly encountered by teachers in their classrooms, as well as clinically relevant symptoms cited by psychologists as evidence of an emerging mental health disorder.

For the current study, one internalizing disorder and one externalizing disorder were selected. Research has shown that separation anxiety disorder (SAD) and oppositional defiance disorder (ODD) are the most prevalent internalizing and externalizing disorders among children in primary school grades (Splett et al., 2019). As primary school teachers and children are the populations of interest, SAD and ODD were selected as the mental health disorders used to design the vignettes. Furthermore, the vignettes described children between the ages of 6 and 11 to represent that age range of children in primary Grades 1 to 6. The inclusion of both an externalizing and internalizing disorder within the vignette measure was done to determine if the ability of participants to discriminate between typical and atypical behaviors was significantly influenced by the type of pathology described, as has been established in prior research (Splett et al., 2019; J. Green et al., 2018).

To account for potential main effects of pathology type and gender, as well as interactions between these variables, as seen in previous research (M. Green et al., 1996; J. G. Green et al., 2018; Loades & Mastroyannopoulou, 2010), two versions of each vignette were created with only the names and gender pronouns differing between them. Gender and pathology combinations were arranged into two vignette measure conditions. Condition one (the congruent condition) contained vignettes describing boys presenting with ODD and girls presenting with SAD. In condition two (the incongruent condition), participants read vignettes describing boys presenting with SAD and girls presenting with ODD. Altogether, twelve unique vignettes were created with two combinations of gender and pathology for each of three levels of behavioral severity (see Supplemental Appendix 1 for vignette measure materials). Before beginning the vignette measure, participants were assigned to either condition using a pseudo-randomization sorting algorithm. The order in which vignettes were presented was also randomized to account for order effects.

To create a nuanced dimensional depiction of ODD and SAD at varying levels of symptoms severity, three distinct depictions were created for each pathology. Building upon information collected from the ABCs of Mental Health, a free resource for teachers

and parents provided by the Hincks-Dell Crest Centre at Sick Kids Hospital, symptoms were described according to three levels of severity, mild, moderate, and severe. Mild severity vignettes described behaviors within a normal range of expectations for children of that age. Behaviors within this range may be problematic for academic and social functioning but are transient and not so intense as to be beyond the ability of a teacher to adequately address. Moderate severity vignettes described specific behaviors that are beyond normal expectations for children of that age and warrant further consideration. Moderate severity vignettes, and the behaviors depicted in them, are more indicative of emerging mental health disorders and include more intense and frequent problematic behaviors occurring over longer time spans (e.g., months compared to weeks in the mild vignettes). Severe severity vignettes described behaviors that are significantly outside of a typical developmental range and warrant referral to a mental health specialist. These are problematic behaviors that occur frequently in the classroom, are intense enough to create significant impairment in multiple domains and persist for many weeks or months. One of the key aspects of vignettes in the severe category is the inability of the teacher to adequately address the behaviors depicted without support.

In addition to varying degrees of problematic behaviors described across the three levels of vignette severity, a thematic analysis was conducted to ensure a standardization of content between and across severity levels. Identifying and standardizing thematic components provided a complex assortment of information participants could integrate and use to support their judgments of behavioral severity. Through this, participants were presented with the same kinds of diagnostic information that would typically be available to them in their real-world classrooms such as length of observation, frequency, intensity, and duration of the behaviors, and external factors (e.g., family stressors).

Procedure

Participants accessed the online survey via an email with embedded hyperlink. The survey consisted of several demographic questions, the novel vignette measure, and the Mental Health Literacy Scale created by O'Connor and Casey (2015). For the purpose of this study, only data pertaining to participant's responses on the vignette measure is examined.

Participants were presented with six vignettes, one at a time. After reading each vignette, participants rated the behavior of the child depicted in the vignette on a scale from 1 to 5. A rating of one would indicate that the behaviors depicted are within a normal developmental range and not indicative of an emerging mental health disorder. A rating of 5 would indicate that the behaviors depicted is outside a normal developmental range, beyond the ability of the teacher to properly address, and requires referral to a mental health specialist.

Participants' discriminatory accuracy across the vignette measure was calculated as the difference between their behavioral severity rating (1–5) and the underlying three-point severity structure. In order to account for potential biases in extreme responding (e.g., a proclivity to choose a rating of 1 or 5 only), participants' ratings were transformed to fit with the three-point severity structure of the vignettes. See Figure 1 below for a diagram of how these transformations were conducted:

Vignette Severity	Mild		Moderate	Severe	
Raw Likert Scale Response	1	2	3	4	5
Transformed Score Bracketing	1		2	3	

Figure 1. Vignette Measure Transform Diagram.

A rating of 1 or 2 was treated as a mild severity rating (1) and a rating of 4 or 5 was treated as a severe rating (3). Only a rating of 3 on our Likert scale was treated as a moderate rating (2). Using score bracketing in this manner allowed the researchers to account for potential central tendency biases that may have influenced participant responses on the Likert scale (Douven, 2018).

Accuracy scores on the vignette measure ranged from 0 to 10. Each participant responded to six vignettes. Discriminatory accuracy was calculated as the absolute difference between their rating and the intended severity level of the vignette. For example, a participant rating a mild severity vignette as a 3 on the Likert scale would commit a rating error of 1 (*See below for the VSES formula*). Across all six vignettes, the maximum vignette sum error score (VSES) was 10. A VSES of 10 would represent a maximally erroneous rating, in which each mild vignette was rated as severe, each severe vignette was rated as mild, and each moderate vignette was rated as either mild or severe. A VSES of 0 would represent perfect discrimination, in which the participant was able to rate the severity of the vignette in accordance with the intended behavioral severity depiction.

$$VSES = \sum (| X_i - Y_i |)$$

X_i =The participants behavioral severity rating on a given vignette

Y_i =The intended severity coding for that vignette

Quantizing participant’s discrimination accuracy along a continuous as a sum total of discrimination errors (VSES), allows for statistical analysis using continuous scales and therefore expands upon the utility of previous dichotomized vignette measures. For the purpose of this study, VSES scores were used to examine difference in the congruent and incongruent gender/pathology vignette conditions for the purpose of collapsing across the sample data points.

Results

Ratings and Discrimination Accuracy

Participants’ average behavioral severity ratings across the various vignettes were calculated and can be seen in the table below. The ratings presented below represent raw score ratings made by participants on the 5-point Likert response scale (Table 1).

Table 1. Mean Vignette Severity Ratings by Severity, Pathology, and Gender.

Vignette	Mild	Moderate	Severe
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Externalizing			
Male	1.98 (0.99)	2.64 (0.93)	3.93 (0.97)
Female	2 (1.07)	2.88 (0.85)	3.91 (0.97)
Internalizing			
Male, <i>n</i> = 58	1.93 (1.16)	2.84 (1.11)	4.02 (1.01)
Female	1.56 (0.73)	2.69 (1.06)	3.98 (0.96)

Table 2. Accuracy of Participant' Severity Ratings by Severity, Pathology, and Gender.

Vignette	Mild	Moderate	Severe
	Accuracy (%)	Accuracy (%)	Accuracy (%)
Externalizing			
Male	68.9	37.9	74.1
Female	76.7	55.8	65.1
Internalizing			
Male	69.7	39.5	72.0
Female	89.6	25.8	70.6

Severity ratings across the vignettes indicate that regardless of pathology and gender combinations, participants' severity ratings increase according to the underlying severity structure. Such that mild coded vignettes produce the lowest severity ratings and severe coded vignettes produce the highest severity ratings. Further examination of these severity ratings via modeling via a binomial logistic regression is discussed below.

The accuracy of participants' ratings per vignette are presented below in Table 2. The accuracy estimates below are based on per vignette ratings. Accuracy per vignette was measured according to participants' Likert scale ratings matching the intended severity coding (see Figure 1).

Accuracy estimates indicate that participants were more accurate when rating mild or severe vignettes. Lower accuracy on the moderate vignettes may represent the potential for participants to commit rating errors in two directions, either under or over rating.

Comparing Discrimination Accuracy Across Gender/Pathology Conditions

An independent samples *t*-test was conducted to determine if significant differences in vignette discrimination scores were observed between the two gender/pathology vignette conditions.

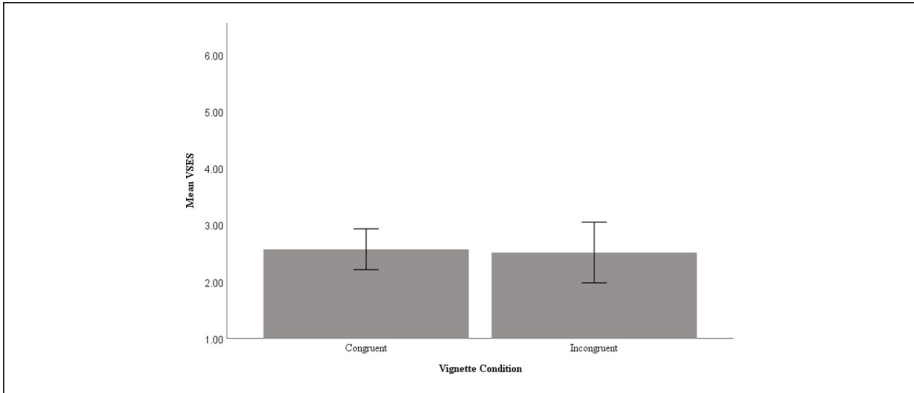


Figure 2. Comparison of mean vignette sum error scores by vignette measure condition. Note. The congruent vignette condition consisted of male/externalizing vignettes and female/internalizing vignettes. The incongruent vignette condition consisted of male/internalizing and female/externalizing vignettes.

There were 58 participants in the congruent vignette condition and 43 participants in the incongruent vignette condition. A Welch *t*-test was run to determine if there were differences in participants' vignette discrimination performance between the congruent and incongruent conditions due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances ($p = .034$). Outliers were present in the data, as assessed by inspection of a boxplot, and vignette sum error scores for each condition were found to be non-normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$). Participants in the congruent condition made slightly more discriminatory errors ($M = 2.57$, $SD = 1.37$) as compared to those in the incongruent condition ($M = 2.50$, $SD = 1.70$), but the difference was not statistically significant (95% CI [-0.58, 0.69]), $t(77.32) = 0.179$, $p = .858$ (Figure 2).

Main Effects and Interactions of Gender, Pathology, and Severity on Discrimination Accuracy

To explore the effect of gender, pathology type, and severity on participants' discrimination accuracy across the vignette measure, a generalized linear model using a binary probability distribution was used. Within this model the dependent variable, per vignette trial discrimination error, was dichotomized as only 4.3% of cases were observed in which participants made a discrimination error of more than 1 point. In other words, out of the 660 vignette ratings made within our data set, only 26 ratings were maximally erroneous (rating a mild vignette severe or vice versa). Therefore, the dependent variable was dichotomized to reflect the probability of a participant either correctly or incorrectly rating the severity of the vignette. Participants' mental health literacy scores on the MHLS (O'Connor & Casey, 2015) were added as a covariate to control for the influence of this variable on their discrimination performance.

Table 3. Test of Model Effects, Binomial Logistics Regression.

Source	Wald chi-square	df	p
Severity	85.600	2	.000
Gender	1.957	1	.162
Pathology	0.000	1	.988
Severity × Gender	5.756	2	.056
Gender × Pathology	0.016	1	.899
Severity × Pathology	6.650	2	.036
Severity × Gender × Pathology	9.385	2	.009

Dependent variable: Vignette Discrimination.

Table 4. Significant Pairwise Comparisons of Estimated Marginal Means Between Severity Levels.

(I) Behavioral severity of each vignette	(J) Behavioral severity of each vignette	Mean difference (I-J)	Std. error	df	Bonferroni sig.	95% Wald confidence interval for difference	
						Lower	Upper
Mild	Moderate	0.38 ^a	0.048	1	0.000	0.27	0.49
	Severe	0.07	0.053	1	0.619	-0.06	0.19
Moderate	Mild	-0.38 ^a	0.048	1	0.000	-0.49	-0.27
	Severe	-0.31 ^a	0.041	1	0.000	-0.41	-0.21
Severe	Mild	-0.07	0.053	1	0.619	-0.19	0.06
	Moderate	0.31 ^a	0.041	1	0.000	0.21	0.41

^aThe mean difference is significant at the 0.05 level.

Test of model effects indicate a significant main of effect of severity $X^2(2, 86.60)$, $p < .001$. A significant interaction was also observed between severity and pathology $X^2(2, 6.65)$, $p = .036$, as well as severity, pathology, and gender $X^2(2, 9.385)$, $p = .009$ (Table 3).

Examination of pairwise comparisons using estimated marginal means (see Supplemental Appendix 2) revealed that differences in the likelihood of committing rating errors were primarily driven by differences in the underlying severity of the vignettes, rather than the gender or pathology described in the vignettes.

Differences in rating error likelihood were only observed between moderate/mild and moderate/severe coded vignettes. This means that the probability of making an error was only greater when participants rated moderate severity vignettes as compared to mild and high severity vignettes. Based on the structure of the vignette measure this makes sense, given that when presented with a moderate severity vignette participants had the potential to make rating errors in two directions, whereas mild vignettes could only be incorrectly rated in the positive direction (over rated) and severe vignettes could only be incorrectly rated in the negative direction (under rated; Table 4).

Discussion

In line with previous research employing multiple vignettes depicting clinical and subclinical elevations in behaviors indicative of mental health disorders (M. Green et al., 2018), participants' ratings appeared to be primarily driven by the underlying severity construct. Vignettes depicting behaviors within a typical developmental range were rated least serious, moderate severity vignettes were rated more serious than the mild vignettes, and high severity vignettes were rated as being the most serious. Examination of rating accuracy by vignette type reveals low rates of accuracy for the moderate severity vignettes, and higher rates of accuracy for the mild and severe level vignettes. Although participants overall made few discrimination errors on average across the vignette measure, accuracy of problem identification across the severe and mild vignettes was lower than estimates from previous research (Splett et al., 2019). Across the severe externalizing vignettes, participants correctly rated them 69.6% of the time. Participants were similarly accurate across the severe internalizing vignettes, rating them accurately 71.3% of the time. Previous research by Splett et al. (2019) using a dichotomized outcome variable for level of concern found accuracy rates of 100 and 98% for their participants' identification of severe externalizing and internalizing problems respectively. Results of the current study suggest that when vignette measures are designed to allow participants to rate various behaviors on a continuous scale, their accuracy of discrimination is lower than when they are asked to simply indicate the presence of absence of a problem. Our accuracy estimates are more similar to those derived through classroom nomination methods (Cunningham & Suldo, 2014). This alignment in accuracy estimates with classroom nomination methods, suggests our novel vignette measure provides a more ecologically valid estimate of the discriminatory identification decisions teachers make in their real world classrooms, as compared to previous vignette measure design studies.

Significant main effects or interactions for gender or pathology on discrimination accuracy were not found, in contrast to previous studies (Loades & Mastroyannopoulou, 2010). The absence of such gender and pathology effects and the presence of significant differences in ratings based primarily on the severity of the behaviors described, may indicate that this novel vignette measure is less vulnerable to biases in ratings based of the gender of the child and the type of pathology depicted within the vignettes as previously observed (M. Green et al., 1996; Loades & Mastroyannopoulou, 2010).

The discrimination errors that participants did commit appeared to be driven primarily by the severity of the vignettes. There was no evidence in our data that error rates within severity levels were significantly influenced by the gender or pathology of the depicted students. Participants were only more prone to making errors when rating moderate severity vignettes. The inclusion of a three-tiered severity construct within our vignette measure is to the knowledge of the researchers, a novel addition. The inclusion of a dimensional severity structure may also explain the absence of observed gender and pathology effects. Providing dimensional nuance within the depictions of behaviors and presenting more vignettes than typically done in previous studies, may have encouraged participants to integrate the diagnostic information in such a way that their judgments were less influenced by considerations of gender or pathology expressions.

The predominance of the underlying severity structure in its influence on respondent ratings is encouraging from a psychometric standpoint, as it suggests that the instrument is capturing discriminatory judgments based primarily on behavioral severity, as intended. Furthermore, the lower levels of discriminatory accuracy observed in this studies' novel vignette measures are more closely aligned with accuracy rates seen through more robust measures of teacher identification accuracy via classroom nominations (Cunningham & Suldo, 2014). Capturing the discriminatory accuracy of teachers in differentiating between the severity of presented behaviors within our vignette measure may provide a quick, cheap, and easy means by which to assess educators' ability to differentiate between typical and indicative student behaviors.

In conclusion, the participants in the current study were less accurate in their ratings of behavioral severity than found in previous studies, and when they did commit rating errors, they appeared to be contingent upon the underlying severity structure rather than gender or pathology biases.

Limitations and Future Directions

There are several limitations inherent in using a vignette methodology like the one employed in the current study. Although the use of vignettes allowed the researchers to assess participant responses to a standard stimulus, in a real-world context teachers may judge the behaviors of actual students different than those presented in vignettes (Jerolmack & Khan, 2014). Furthermore, the nature of vignettes is such that often difficult to detect behaviors associated with internalizing disorders are made explicit, as with our depictions of separation anxiety. Classroom realities may also render teachers unable to attend to the individual behaviors of all the students in their classrooms. Therefore, the unambiguous explication of behaviors that are typically hard to detect, limits the extrapolation of these results to real world classroom contexts. Indeed, research suggests that externalizing behaviors are more likely to be identified by teachers due to their more salient and visible nature (Walker et al., 1984).

This study evaluated some of the most common forms of internalizing (separation anxiety) and externalizing (oppositional defiance disorder) disorders in primary school age children. This narrow band of disorder types limits the extension of this studies results to estimating the general discriminatory ability of teachers to differentiate between behaviors indicate of other forms of psychopathology. Nor do these results elucidate how teachers would rate the behaviors of students presenting with behaviors indicative of multiple comorbid psychopathologies.

Further validation of this novel vignette measure should be conducted to establish its validity and reliability. Comparing discriminatory accuracy between our novel vignette measure and accuracy of student identification through classroom nominations by teachers may provide a means of establishing convergent validity.

Relevance to the Practice of School Psychology

Help-seeking pathway models indicate that for children who experience mental health problems, teacher recognition of the problem is typically the first step in seeking help

(Bussing et al., 2003). The educational context is unique in its ability to elicit student behaviors indicative of psychopathologies, given the unique demands that classrooms place on students (Narad et al., 2015). Furthermore, the presentation of behaviors indicative of emerging mental health problems may be more salient to teachers in a context that allows for comparison to the norm, as would be the case in a classroom of many children. A report by the Standing Senate Committee on Social Affairs, Science and Technology chaired by Kirby and Keon (2006), identified the untapped potential of schools to facilitate mental health prevention and intervention for children and youth.

Early detection of emerging mental health disorders in students is critical to ensuring positive long-term outcomes. However, the provision of mental health resources within educational contexts is also limited by systematic constraints that exist with educational and mental health systems in Canada. A survey conducted in 2014 by Children's Mental Health Ontario revealed that 28,000 children were on active wait-lists for mental health services with average wait times of 2 to 3 months for therapy or intensive treatment (Children's Mental Health Ontario, 2020). Given service seeking often occurs after psychological assessments have been completed, the time between a child experiencing a mental health disorder and being identified, referred for assessment, and then referred to treatment is likely considerably longer.

While students wait to be assessed, the likelihood of their disorders developing into more intractable complaints is increased, and thereby limits the potential for full recovery and optimal beneficial long-term outcomes. Therefore, it is important that school board referrals by teachers accurately capture students experiencing the onset of mental health disorders. Improving the Addressing mental health problems in students early on, before these problems have developed into mental disorders, would consequently also relieve pressure on already taxed mental health resources and supports (Njie et al., 2017).

The current study attempted to explore the utility of a novel vignette measure in assessing the discriminatory accuracy of teachers in their judgment of behaviors indicative of emerging mental health disorders in students. Using this measure, school boards could quickly and accurately assess the accuracy with which their teachers are making their mental health referral decisions. In doing so, gaps in mental health literacy that contribute to the misallocation of school board mental health resources (e.g., not over referring or under referring) could be identified and addressed at a systems level through professional development or other training initiatives. It is the hope of the researchers that this measure be implemented by school boards to identify inequities in school based help-seeking pathways that inhibit the timely and judicious provision of mental health resources for struggling students.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Damian Page  <https://orcid.org/0000-0002-5878-2151>

Supplemental Material

Supplemental material for this article is available online.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*.
- Bussing, R., Zima, B. T., Gary, F. A., & Garvan, C. W. (2003). Barriers to detection, help-seeking, and service use for children with ADHD symptoms. *The Journal of Behavioral Health Services & Research, 30*(2), 176–189.
- Children's Mental Health Ontario. (2020). *Kids can't wait: 2020 report on wait lists and wait time for child and youth mental health care in Ontario*. <https://cmho.org/wp-content>
- Cunningham, J. M., & Suldo, S. M. (2014). Accuracy of teachers in identifying elementary school students who report at-risk levels of anxiety and depression. *School Mental Health, 6*(4), 237–250.
- Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin & Review, 25*(3), 1203–1211.
- Green, J. G., Guzmán, J., Didaskalou, E., Harbaugh, A. G., Segal, N., & LaBillois, J. (2018). Teacher identification of student emotional and behavioral problems and provision of early supports: A vignette-based study. *Journal of Emotional and Behavioral Disorders, 26*(4), 225–242.
- Green, M. T., Clopton, J. R., & Pope, A. W. (1996). Understanding gender differences in referral of children to mental health services. *Journal of Emotional and Behavioral Disorders, 4*(3), 182–190.
- Jerolmack, C., & Khan, S. (2014). Talk is cheap: Ethnography and the attitudinal fallacy. *Sociological Methods & Research, 43*(2), 178–209.
- Jorm, A. F., Korten, A. E., Jacomb, P. A., Christensen, H., Rodgers, B., & Pollitt, P. (1997). "Mental health literacy": A survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Medical Journal of Australia, 166*(4), 182–186.
- Kirby, M. J., & Keon, W. J. (2006). *Out of the shadows at last: Highlights and recommendations* (Final report). Standing Senate Committee on Social Affairs, Science, and Technology.
- Loades, M. E., & Mastroyannopoulou, K. (2010). Teachers' recognition of children's mental health problems. *Child and Adolescent Mental Health, 15*(3), 150–156.
- Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: Developmental epidemiology. *Journal of Child Psychology and Psychiatry, 45*(3), 609–621.
- Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., Simon, J. O., & Epstein, J. N. (2015). Parent–teacher agreement on ADHD symptoms across development. *Psychological Assessment, 27*(1), 239.
- Njie, A., Shea, S., & Williams, M. (2017) *Turning point. Interim report of the commission on inclusive education*. Copyright Commission on Inclusive Education. https://www.inclusivedns.ca/wp-content/uploads/2017/06/CIE_Report_English.pdf

- O'Connor, M., & Casey, L. (2015). The mental health literacy scale (MHLS): A new scale-based measure of mental health literacy. *Psychiatry Research, 229*(1–2), 511–516.
- Santa Lucia, R. C., Gesten, E., Rendina-Gobioff, G., Epstein, M., Kaufmann, D., Salcedo, O., & Gadd, R. (2000). Children's school adjustment: A developmental transactional systems perspective. *Journal of Applied Developmental Psychology, 21*(4), 429–446.
- Splett, J. W., Garzona, M., Gibson, N., Wojtalewicz, D., Raborn, A., & Reinke, W. M. (2019). Teacher recognition, concern, and referral of children's internalizing and externalizing behavior problems. *School Mental Health, 11*(2), 228–239.
- Stewart, D. E. (2003). "Physical symptoms of depression: Unmet needs in special populations." *Journal of Clinical Psychiatry, 64*(suppl 7), 12–16.
- Walker, E., Bettes, B., & Ceci, S. J. (1984). Teachers' assumptions regarding the severity, causes, and outcomes of behavioral problems in preschoolers: Implications for referral. *Journal of Consulting and Clinical Psychology, 52*(5), 899.

Author Biographies

Damian Page is a clinical and school psychology doctoral student at the Ontario Institute for Studies in Education (OISE) at the University of Toronto. Damian's research work focuses on improving help-seeking pathways for students with mental health needs through mental health literacy.

Dr. Todd Cunningham is a clinical and school psychologist and Assistant Professor at the University of Toronto / OISE. Dr. Cunningham's innovative research investigates the integration of assistive technology and learning strategies for children with learning difficulties due to a variety of reasons including mental health challenges.