

# Article Graded Response Modeling of the DESSA's Self- and Social Awareness Subscales in a Special Needs Sample

David Adams<sup>1</sup> and Gregory Lawrence Farmer<sup>2,\*</sup>



<sup>2</sup> Graduate School of Social Service, Fordham University, New York, NY 10458, USA

\* Correspondence: farmer@fordham.edu

Abstract: Addressing the interpersonal challenges that students with Emotional Disturbance (ED) classifications experience is critical to their success in and outside the school setting. Improving their self- and social awareness will strengthen their ability to navigate social relationships in and outside school. The planning and evaluation of interventions targeting the development of self- and social awareness requires psychometrically sound assessment instruments. Using the Graded Response Item Response Theory (GRM) approach, this study provides evidence of the reliability of the self- and social awareness subscale of the Devereux Student Strengths Assessment (DESSA) among students with an ED classification. The study participants were a sample of 449 youth attending schools serving students classified as emotionally disturbed in self-contained and inclusion settings. The examination of GRM item parameters (i.e., item discrimination and difficulty) and the information curves (i.e., test and item information curves) provides evidence of the reliability of DESSA's self- and social awareness subscales across a broad range of students' levels of self- and social awareness.

**Keywords:** graded response model; item response theory; social and emotional learning; special education; emotional disturbance

# 1. Introduction

For those students with an emotional disturbance (ED) classification, the acquisition of social-emotional learning competencies (SEC) is critical to their academic and interpersonal success [1–3]. Children with learning disabilities struggle to develop their SEC. For example, they have difficulty developing the ability to recognize emotions in themselves and others, i.e., self- and social awareness, the ability to regulate and manage emotions, i.e., self management, and to set positive and realistic goals, i.e., goal-directed behavior [1]. Selfawareness represents the ability of individuals to identify their values and feelings, along with having a realistic understanding of their strengths and limitations [4,5]. The ability to take another's perspective and understand social group norms is reflected in an individual's level of social awareness [4,6]. Both self- and social awareness are essential in order to successfully navigate the social environment in and outside the school setting [7]. Self- and social awareness help to contribute to the individual's ability to engage with their social and physical environment [8]. Research provides evidence that self-awareness contributes to an individual's ability to cope with challenges [9]. Social awareness has been found to contribute to the development of self-management, responsible decision-making, and relationship skills [8].

There is overwhelming evidence that well-designed and implemented SEL programs improve students' academic performance, increase prosocial behaviors, and reduce conduct and internalization problems [10,11]. The evidence of the value of SEL programs to improve students' academic development has resulted in nearly every state in the United States incorporating SEL into their educational standards [12]. The evidence supporting the effectiveness of SEL programs is less extensive for students in the special needs population



Citation: Adams, D.; Farmer, G.L. Graded Response Modeling of the DESSA's Self- and Social Awareness Subscales in a Special Needs Sample. *Youth* 2022, *2*, 98–112. https:// doi.org/10.3390/youth2020008

Academic Editor: Todd Michael Franke

Received: 18 January 2022 Accepted: 22 March 2022 Published: 25 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). than it is for those in the general education population [13,14]. In a recent meta-analysis of 166 studies evaluating the effectiveness of universal SEL programs, only 19 specifically included students with disabilities [13]. A second meta-analysis study found the absence of SEL intervention studies focusing on children with learning disabilities [14]. Together, those two systematic reviews indicate the need for greater attention on the impact of SEL programming on students with learning disabilities. A review of the measures used to evaluate SEL programs' effectiveness highlights the importance of measuring outcomes which are specific to the focus of the SEL intervention [15]. The continued development and evaluation of SEL programs that improve the SEL competencies of students with learning disabilities depend on the use of measures of SEL that have sound psychometric properties for that population of students. For those with a learning disability resulting in an ED classification, the promotion of their self- and social awareness competency is critical for their social and academic success. The evaluation of SEL programs focused on those competencies requires SEL measures with information about psychometric properties with this population of students.

The Devereux Student Strengths Assessment (DESSA) is a widely used strength-based measure of SEL competencies [13,16]. The DESSA is grounded in the Collaborative for Academic, Social, and Emotional Learning (CASEL) conceptualization of SEL [17]. It focuses on the identification of individual children's social and emotional strengths and needs while also providing classroom or school-wide profiles that can guide universal prevention strategies [18]. In a review of 73 measures used to assess the SEL of middle school students, the DESSA's ability to provide SEL profiles at the student, classroom, and school levels is a strength compared to other SEL measures [18]. The DESSA allows for the examination of individual students, classrooms and grade levels, strengths in the five domains of SEL competencies (i.e., Self- and Social Awareness; Self-Management, Responsible Decision-Making and Relationship Skills), and Optimistic Thinking. This supports efforts to measure outcomes which are specific to the competencies that are the focus of an intervention, thereby retaining a nuanced understanding of the intervention. For those SEL focusing on the self-and social awareness of children with an ED classification, the DESSA could provide a valid strength-based assessment of those competencies.

There is evidence that the internal consistency reliability coefficient for the teacher version of the DESSA among students in general education programs ranges from 0.87 to 0.93 [16]. The internal consistency subscale ranges from 0.92 (Self-Management Subscale) to 0.92 (Personal Responsibility) for teachers, and parents range from 0.86 (Self-Management subscale) to 0.98 (Personal Responsibility) [16]. The overall social-emotional composite score for a parent was 0.98, and the teacher was 0.99 [16]. The one-week test-retest for the parents' version of the DESSA is 0.90 [17]. The study of the convergence between parent and teacher ratings found a correlation between the subscales ranging from 0.49 to 0.78 [17]. The correlation between the DESSA and two widely used measures of adaptive and maladaptive behaviors in school-age children, the Behavioral and Emotional Rating Scale 2nd edition (BERS-2) and the Behavior Assessment System for Children 2nd edition (BASC-2), supports the concurrent validity of the measure [17]. There is currently limited research on the psychometric properties of DESSA with special needs populations. The evaluation of the psychometric properties of DESSA students with special needs is limited to the examination of the difference in the DESSA composite SEL score [16,19]. An understanding of the reliability across the range of Self- and Social Awareness present in heterogeneous groups of youth with special education classifications like ED is unavailable for DESSA subscales. The psychometric analysis of the two subscales based on Item Response Theory (IRT) will allow for a detailed assessment of the conditional reliability of the scales [20]. Conditional reliability refers to the way in which the reliability of a measure changes depending on the level of the construct being measured [21]. It will also examine how each item contributes to the scale's reliability [20].

The Graded Response Model (GRM) that is the focus of this study is one of several Polytomous IRT Models [22,23]. The GRM extends the binary IRT latent psychometric

modeling technique that focuses on an item-level analysis of a measure [24]. Like factor analysis, GRM provides evidence of a measure's construct validity and reliability. GRM includes information on how each item or indicator of the construct being measured contributes to the measure's ability to differentiate among individuals expressing different competency levels [25]. GRM, like all IRT models, can distinguish person and item parameters. The person parameter is referred to as "ability" or a latent trait using the Greek letter theta ( $\theta$ ). In the context of measures of social constructs, "ability" is the level of the construct demonstrated by the individual, for example, the level of social awareness. The GRM model evaluates the relationship between the latent variable and the item response using one or more item parameters [26,27]. The GRM estimated in this study extends the two-parameter binary IRT model (2-PL) used for dichotomous scales [26,28]. Like the 2-PL IRT model, two item parameters, item discrimination and difficulty, are used to characterize the relationship of the items on the underlying latent variable being assessed.

Discrimination parameters evaluate how well an item discriminates between individuals scoring high and low on a latent construct or "ability" [29]. Items can vary in the range of the levels of the latent construct; they can discriminate between high and low scores or ratings. Item difficulty describes how high on a latent trait (e.g., what level of self-awareness) a student typically has before the teacher will endorse a particular rating of an item (e.g., rate a particular behavior as occurring frequently). The item difficulty parameters reflect the points at which a respondent with a given latent trait has an equal probability (50:50) of endorsing or selecting a particular rating associated with an item [20]. For example, the level of self-awareness that a student demonstrates would result in a teacher rating the item 'Describe how he/she was feeling?' a '4', or 'Very Frequently'. The total number of response categories leads to the threshold estimated for measures using ordered rating scales [22]. Because there are five response categories (0 = Never to 4 = Very Frequently), four-item difficulty parameters were estimated corresponding to each rating category threshold (b<sub>1</sub>, b<sub>2</sub>, b<sub>3</sub>, and b<sub>4</sub>) [28].

Using GRM, several research questions related to the DESSA Self- and Social Awareness subscales' psychometric properties in a sample of students with a special education classification of Emotional Disturbance (ED) were examined. Because the DESSA is a copyrighted measure, the exact items are not presented; see Table 1 to illustrate the items. The following questions were examined:

**Table 1.** Self-awareness and social awareness subscales' descriptive statistics and factor analysis parameters (n = 437).

	Mean +	SD	Item Total Correlation	F1(h <sup>2</sup> ) ++
Self-Awareness ( $\alpha = 0.89$ )				
Self 1. Accurately talks about life events?	1.80	0.86	0.62	0.70 (0.49)
Self 2. Teaches others to do something?	1.63	0.98	0.70	0.82 (0.68)
Self 3. Ask questions to clarify	1.71	1.00	0.73	0.87 (0.76)
understanding?				
Self 4. Is aware of personal strengths?	1.62	0.97	0.78	0.90 (0.81)
Self 5. Ask for feedback?	1.43	0.98	0.69	0.85 (0.73)
Self 6. Describes feelings?	1.76	0.92	0.65	0.71 (0.50)
Self 7. Give an opinion when asked?	2.18	0.92	0.59	0.71 (0.50)
Social Awareness ( $\alpha = 0.93$ )				
Socaw 1. Cope with insults?	1.20	0.93	0.65	0.74 (0.56)
Socaw 2. Get along with others?	1.95	0.90	0.75	0.86 (0.74)
Socaw 3. Act respectfully when competing?	1.89	0.90	0.76	0.88 (0.77)
Socaw 4. Respect another person's opinion?	1.77	0.89	0.81	0. 91 (0.82)
Socaw 5. Contribute to group efforts?	1.90	0.95	0.65	0.75 (0.56)
Socaw 6. Resolve a disagreement?	1.27	0.89	0.69	0.80 (0.66)
Socaw 7. Share with others?	1.94	0.94	0.73	0.81 (0.66)
Socaw 8. Cooperate with peers or siblings?	1.84	0.85	0.80	0.91 (0.82)
Socaw 9. Forgive others who hurt her/him?	1.74	0.96	0.74	0.82 (0.67)

Note: + Theoretical range of the means '0' = Never to '4' = Very Frequently. ++ Factor analysis parameters:  $F1 = Factor loading; h^2 = Communalities.$ 

"Is the level of precision of the subscales sufficient across the levels of the traits being assessed?" The item parameters (item discrimination and item difficulty thresholds) provide information about the areas of the latent trait continuum that is measured with precision. Additionally, information curves visually display the changes in the accuracy of the subscale across levels of the construct being assessed. Test information is related to reliability of information values for approximately 2, 4, and 8, corresponding to reliability coefficients of 0.70, 0.80, and 0.90, respectively [30].

"Are there subscale items that contribute more to the quality of subscales precision than other items?" The examination of the item information curves provides evidence of the contribution of each item to the precision of the scale.

## 2. Materials and Methods

# 2.1. Participants

This study was a secondary analysis of data gathered as part of the assessment component of the program of supports developed in a large urban district in the northeast of the United States that serves students with special needs. All of the ethical guidelines for the protection of human subjects were followed. The range of ages in the sample was five to fourteen years old (N = 437, M = 10.16, SD = 2.34). The sample included data from five specialized elementary and middle schools serving students classified as ED in a self-contained setting. These schools used the teacher rating version of the DESSA. The teacher ratings were collected at the beginning and the end of the year in order to evaluate student growth in the social-emotional domain. The sample was 84% male and 16% female, in line with research on gender disparities for students classified as emotionally disturbed [31]. The schools included in the sample had active programs targeting students' social-emotional and behavioral development that included the use of school-wide Positive Behavior Supports and Social–emotional Learning Curricula at the time of the assessment. Additionally, all of the students received counseling services from the school, targeting social-emotional competencies, with counselors and teachers discussing assessment results with students and using them to plan interventions to remediate any skill gaps [32].

#### 2.2. Measures

The teacher ratings were collected in the beginning, and provided the data for this study. The following is a brief description of the self-awareness and social awareness subscales on the teacher version of the DESSA (See Table 2). The seven items making up the self-awareness subscale indicate the extent to which the youth have a realistic understanding of their strengths and opportunities for growth. The subscale also reflects self-improvement. The nine items contained on the social awareness subscale indicate the extent of the youth's capacity to interact with others in a way that demonstrates respect, cooperation, and tolerance.

Table 2. Graded response model IRT parameters.

	а	$b_1$	$b_2$	$b_3$	$b_4$	$b_4 - b_1$
Self-Awareness +						
Self 1. Accurately talks about life events?	1.818	-2.125	-0.534	1.215	2.861	4.986
Self 2. Teaches others to do something?	2.334	-1.339	-0.228	1.203	2.345	3.684
Self 3. Ask questions to clarify understanding?	2.778	-1.319	-0.262	0.923	2.249	3.568
Self 4. Is aware of personal strengths?	3.466	-1.243	-0.124	1.021	2.226	3.469
Self 5. Ask for feedback?	2.534	-1.064	0.084	1.397	2.322	3.386
Self 6. Describes feelings?	1.887	-1.872	-0.396	1.155	2.594	4.466
Self 7. Give an opinion when asked?	1.681	-2.854	-1.050	0.633	2.045	4.899
Social Awareness ++						
Socaw 1. Cope with insults?	1.847	-0.959	0.463	1.829	3.450	4.409
Socaw 2. Get along with others?	2.661	-1.979	-0.628	0.754	2.251	4.230
Socaw 3. Act respectfully when competing?	3.043	-1.868	-0.579	0.926	1.950	3.818
Socaw 4. Respect another person's opinion?	3.589	-1.657	-0.372	1.022	2.118	3.775
Socaw 5. Contribute to group efforts?	1.874	-2.058	-0.583	0.928	2.329	4.388
Socaw 6. Resolve a disagreement?	2.176	-1.042	0.302	1.954	3.001	4.043
Socaw 7. Share with others?	2.570	-1.836	-0.624	0.741	2.160	3.995
Socaw 8. Cooperate with peers or siblings?	3.448	-1.771	-0.523	0.959	2.267	4.039
Socaw 9. Forgive others who hurt her/him?	2.417	-1.591	-0.315	1.011	2.377	3.968

Note: + RMSEA = 0.072, 95 CI [0.061, 0.083] SRMSR = 0.054, CFI = 0.974; TLI = 0.961. ++ RMSEA = 0.074, 95 CI [0.064, 0.082] SRMSR = 0.037, CFI = 0.976; TLI = 0.982.

#### 2.3. Analysis Strategies

Descriptive and bivariate analyses were carried out using SPSS 25 [33]. The assessment of the factor structure of the subscales was carried out using Factor 10.10.03 [34]. The graded response model was estimated using the R programming environment with the multidimensional item response theory package (MIRT), using the full-information maximum likelihood (FIML) estimator [35,36]. The FIML was found to yield good estimates with small sample sizes, and with non-normal distributions [37,38].

Graded response model: The graded response model was developed to evaluate measures using ordinal responses, such as Likert-type scales [26,39]. While a comprehensive discussion of graded response models is beyond the scope of this paper, three critical assumptions will be discussed, i.e., monotonicity, unidimensionality, and local independence [40]. Those who want more information on the family of graded response models [27,39,41] are excellent sources.

Monotonicity: IRT uses mathematical models to describe the relationship between an individual's ability or level on a latent variable (theta, symbolized as  $\theta$ ) and how they respond to an item [41]. When the response categories are polytomous (i.e., ordinal rating scales)—for example, when a Likert scale is used—the item and ability relationship are visually described by the category response curve [CRC; 24]. Monotonicity assumes a monotonically increasing relationship between the ability to be assessed by an item and the probability of responding to higher response categories as the latent trait increases. The CRC is the likelihood of endorsing a particular response category across the ability level. The examination of the Mokken's scalability coefficient H for the items and total subscale was evaluated in order to provide evidence of the monotonicity of the subscales. The Mokken scaling in the R package estimated the scalability coefficient H indexes (version 2.8.4) [42,43]. When the scalability coefficient H per item is greater than or equal to 0.30 and 0.50 for the total scale, support for monotonicity of the scale is found [43].

Dimensionality: The assessment of the dimensionality of the subscales and a direct evaluation of the local independence of the items is essential to the accurate estimation of the graded response model. The number of dimensions or factors the measure has will determine whether a unidimensional or multi-dimensional graded response model needs to be estimated [39]. Exploratory factor analysis (EFA) using parallel analysis (PA) was used to assess the dimensionality of the measure [44-46]. PA is based on the comparison of the eigenvalues of the actual data to those of the Monte Carlo simulated dataset. PA involves the creation of a Monte Carlo simulated dataset with the same numbers of observations and variables as the original data. A correlation matrix is computed from the randomly generated dataset, and then eigenvalues of the correlation matrix are calculated. Scree plots of both the simulated dataset and original are plotted and compared. When the eigenvalues from the random data are larger than the EFA's eigenvalues, you can assume that the factors are mostly random noise. Therefore, the factors that should be retained are those with eigenvalues larger than those from the simulated dataset [47]. In order to further examine the dimensionality of the scales, the Congruence (UniCo) and Explained Common Variance (ECV) indices were also examined. Those indices examine the closeness to unidimensionality of the measure [48]. A value of UniCo larger than 0.95 suggests that data can be treated as essentially unidimensional [48]. A value of ECV larger than 0.85 indicates that data can be treated as essentially unidimensional [48]. the room mean squared error of approximation (RMSEA) and the comparative fit index (CFI) for the onefactor model will also be examined. An RMSEA at or below 0.06 and a CFI at or above 0.95 will indicate an acceptable fit of the one-factor model [49].

Local independence: The local independence assumption implies that the latent variable underlying the measure is the only factor causing items to co-vary [26]. The likelihood ratio statistic ( $G^2$ ) detects local dependence (LD) by evaluating the difference between the observed and model-predicted item responses from each item pair's contingency table [50,51].

Model Fit: Indexes of both the global and item-level model fits were carried out. The global fit of the model was calculated with the M<sub>2</sub> statistic [52]. The associated RMSEA index was also examined, using the cut-off of being less than or equal to 0.06 [53,54]. The fit of each item to the model was evaluated with the  $S-\chi^2$  index [54], and they were considered to fit the GRM if the associated p > 0.01 [28]. Additionally, MIRT implementation estimates the  $S-\chi^2$  index and computes an RMSEA value, indicating the degree of item-to-model fit. An adequate fit is indicated by values less than 0.06.

#### 3. Results

## 3.1. Item Descriptive Statistics

Table 1 contains descriptive statistics for the items making up the self- and social awareness subscales. The correlation matrix is available upon request. Across both subscales, the item–total correlations ranged from 0.59 to 0.81. As indicated by the reliability coefficients, overall, there was a high degree of inter-correlation among the items.

## 3.2. Dimensionality and Local Independence Assumptions

Self-awareness subscale: The results of the parallel analysis based on minimum rank factor analysis provided evidence that the subscales were essentially unidimensional. Evidence from the Bartlett's statistic [T = 1466.701 (df = 21), p < 0.001] and Kaiser-Meyer-Olkin test [KMO = 0.903; Bootstrapped 95% CI KMO = (0.889, 0.923)] support the appropriateness of the correlation matrix for factor analysis. The first factor for the self-awareness subscale accounted for 73.84% of the variance. The cumulative proportion of variance for the first factor was 0.596. These results were also supported by several closeness-to-unidimensionality indices, such as unidimensional congruence (UniCo = 0.989, BC Bootstrap 95% CI (0.978, 0.997). A UniCo greater than 0.95 suggests that the measure is essentially unidimensional. The explained common variance was ECV = 0.906, 95% CI (0.873, 0.942). An ECV larger than 0.85 suggests that the measure is essentially unidimensional. Table 1 displays the factor loading parameters and the communalities (h<sup>2</sup>) for the unidimensional self-awareness factor. The loading ranged from 0.70 to 0.90, with h<sup>2</sup> ranging from 0.49 to 0.81.

The first factor for the social awareness subscale accounted for 63.0% of the variance. Evidence from the Bartlett's statistic [T = 2478.223 (df = 36), p < 0.001] and Kaiser-Meyer-Olkin test [KMO = 0.940; Bootstrapped 95% CI KMO = (0.936, 0.948)] support the appropriateness of the correlation matrix for factor analysis. The cumulative proportion of variance for the first factor was 0.596. These results were also supported by several closeness-to-unidimensionality indices, such as the unidimensional congruence (UniCo = 0.994., BC Bootstrap 95% CI (0.992, 0.997). The UniCo greater than 0.95 suggests that the measure is essentially unidimensional. The explained common variance was ECV = 0.932, 95% CI (0.919, 0.949). The ECV larger than 0.85 suggests that the measure is essentially unidimensional. Table 1 displays the factor loading parameters and the communalities (h<sup>2</sup>) for the unidimensional social awareness factor. The loading ranged from 0.74 to 0.91, with h<sup>2</sup> ranging from 0.56 to 0.82.

For both the self- and social awareness subscales, none of the LD statistics were greater than 10, indicating no excess of covariation among the item responses when  $\theta$  was held constant [42].

#### 3.3. Monotonically Assumption

The Mokken scalability coefficients for both the subscales provided support for the monotonically assumption. For the self-awareness Mokken, the scalability coefficients Hi of the items ranged from 0.540 (SE = 0.041) to 0.607 (SE = 0.026). The H coefficient for the total scale was 0.585 (0.025). For the social awareness scalability coefficients, the Hi of the items ranged from 0.644 (SE = 0.025) to 0.700 (SE = 0.021). The scalability coefficient H coefficient for the total scale was 0.638(0.021).

## 3.4. Model Fit

Global fit: The global assessment for the self-awareness subscale (M2 = 81.169.24, df = 14, p = 0.001, SRMSR = 0.05) was adequate. This was not consistent with the RM-SEA estimate (RMSEA = 0.072). The global assessment for the social awareness subscale (M2 = 122.101, df = 27, p = 0.001, SRMSR = 0.04) was adequate. This was not consistent with RMSEA estimate (RMSEA = 0.074).

Item fit: The adjustment of each item to the model was evaluated with the  $S \cdot \chi^2$  index. Table 3 displays the  $S \cdot \chi^2$  index for each item. All of the *p*-values were above 0.01, indicating good item-level fit. With evidence of good item level fit and two of the three global fit statistics (M2 and SRMSR) indicating adequate fit, the GRM model's estimation proceeds.

		$S-\chi^2$	Р	RMSEA <i>S</i> - $\chi^2$
Self-Awareness	Self1	23.536	0.656	0.055
	Self2	37.511	0.067	0.028
	Self3	22.952	0.464	0.006
	Self4	21.743	0.475	0.022
	Self5	16.554	0.897	0.015
	Self6	37.229	0.114	0.000
	Self7	19.282	0.420	0.040
Social-Awareness	Socaw1	39.440	0.405	0.028
	Socaw2	17.725	0.886	0.000
	Socaw3	21.476	0.805	0.022
	Socaw4	25.410	0.384	0.000
	Socaw5	36.481	0.672	0.011
	Socaw6	43.848	0.079	0.026
	Socaw7	24.290	0.714	0.019
	Socaw8	16.051	0.886	0.000
	Socaw9	32.082	0.463	0.013

**Table 3.** *S*- $\chi^2$  index and RMSEA *S*- $\chi^2$  values.

df = degrees of freedom.

Parameterization: Table 3 presents the graded response IRT parameter estimates for the self- and social awareness subscales. The items for both subscales cover a wide portion of the latent traits, at -2.854 to 2.861 for the self-awareness scale and -2.058 to 3.450 for the social awareness subscale (See Table 3). The difference between the values of these threshold parameters (b4 - b1) can be interpreted as an indicator of the ease with which a teacher may change their rating from one category to another. In other words, if the differences between the b values are small, small differences in the trait would lead to changes in the teacher's ratings. The discrimination parameters' (a) values range from 3.466 to 1.681 for self-awareness and 1.847 to 3.589 for the social awareness items. It has been recommended that item discriminations between  $1.35 \le a \le 1.69$  should be considered high, and those greater than 1.70 should be identified as very high [55]. This indicates the capability of the response categories to distinguish between different levels of the construct being assessed.

Figures 1 and 2 depict the category response curves (CRC) for the self- and social awareness items. The CRCs for the graded response model parameters are the graphical representation of the probability of rating the different options for a given item across the range of assessed traits. The x-axis on the CRC indicates the underlying level of social or emotional capability being measured on a standardized normal scale ( $\theta$ ). Theoretically, the values could range from  $-\infty$  to  $+\infty$ . The CRC depicts clear delineation for each of the item's rating options [(0) Never to (4) Very Frequently], and a relatively steep rise and fall for the probability of occurrence of each response option.



Figure 1. Self-awareness subscale category response curves.



Figure 2. Cont.



Figure 2. Social-awareness subscale category response curves.

Figures 3-6 display the information and conditional standard error of the measurement functions for the overall scale and each item. In IRT, information refers to the ability of an item and a scale (i.e., a composite score) to accurately estimate the level of the construct being measured (i.e., a point on  $\theta$ ) [20]. It is essential to recognize that information is inversely related to the standard error of measurement [20]. As the information associated with an item increases, so does the item's reliability. Figures 3 and 4 show the information (IF) and conditional standard error of measurement (CSEM) function for the self-awareness subscale. The IF is the top (solid blue) curve, and the CSEM is the bottom (red) curve. Over the range of approximately -2 and +2 theta for the self-awareness subscale, the information index was at or above 4, which corresponds to reliability at or above 0.80. This means that the scale demonstrated good reliability between  $\pm$  two standard deviations from the mean. A close look at the information and the conditional standard error of measurement function for individual items (see Figure 4) indicated that Self 2 (Teach another person to do something), Self 3 (Ask questions to clarify understanding), Self 4 (Show awareness of personal strengths), and Self 5 (Ask for feedback) had the higher levels of information than the other items (Self 1: Accurately asks about life events; Self 6: Describes feelings; Self 7: Give an opinion when asked) The marginal reliability for the self-awareness subscale was 0.91. In IRT, the marginal reliability estimates the overall reliability of the measure based on the average conditional standard error [56]. the self-awareness subscale without those items maintained a marginal reliability of 0.91. The self-awareness subscale without those items achieved a marginal reliability of 0.89. Additionally, the information and conditional standard error of measurement functions did not appreciably change.



Figure 3. Self-awareness scale information and conditional standard error of measurement functions.



Figure 4. Self-awareness subscale item information function.



**Figure 5.** Social-awareness subscale information and conditional standard error of measurement function.



Figure 6. Social-awareness subscale item information function.

Figures 5 and 6 show the information and conditional standard error of measurement functions for the overall scale, and for each item for the social awareness subscale. Over the approximate -2.5 and  $+2.5 \theta$ , the information functions were above 4 for both subscales. The functions indicated reliability at or above 0.80 across that range. The information function show that the scale demonstrated good reliability between  $\pm 2.5$  standard deviations from the mean. A close look at the IF (see Figure 6) indicates that Socaw 2 ("Get along with different types of people"), Socaw 3 ("Act respectfully in a game or competition"), Socaw 4 ("Respect another person's opinion"), and Socaw 8 ("Cooperate with peers or siblings") have higher levels of information than the other items. Socaw 1, 5, 6, 7, and 9 have the lowest amount of information. The social awareness subscale without those items maintained a marginal reliability of 0.90. Additionally, the information and conditional standard error of measurement functions did not appreciably change.

### 4. Discussion

Is the level of precision of the subscales sufficient across the levels of the traits being assessed? Across a broad range of self- and social awareness levels, an adequate level of reliability was achieved for this sample of students receiving special education services. Overall, the subscales for both the self- and social awareness provide the most information in the range of  $-2 \le \theta \le +2$ . For both subscales, the level of reliability in the  $-2 \le \theta \le +2$  range is associated with approximate reliability between 0.80 and 0.95. These findings extend the utility of the DESSA as a strength-based assessment of students' capabilities that can be used to support school's planning and monitoring of interventions focused on developing those competencies.

Are there subscale items that contribute more to the quality of subscales' precision than other items? There was evidence that, for the self-awareness subscales, the following items could be eliminated from the subscale without sacrificing the overall amount of test information.

- Self 1: Accurately talks about life events
- Self 6: Describes feelings
- Self 7: Give an opinion when asked

The item information function for those items was lower than that for other items. The examination of the information function for the self-awareness subscale with only the following items indicated an adequate level of information for the subscale:

- Self 2: Teaches others to do something
- Self 3: Ask questions to clarify understanding
- Self 4: Is aware of personal strengths
- Self 5: Ask for feedback

The following items on the social awareness subscale contributed less than the other items to the level of information the scale provides:

- Socaw 1: Cope with insults
- Socaw 5: Contribute to group efforts
- Socaw 6: Resolve a disagreement
- Socaw 7: Share with others
- Socaw 9: Forgive others who hurt her/him

The Social Awareness subscale with only the following items demonstrated a high level of information:

- Socaw 2: Get along with others
- Socaw 3: Act respectfully when competing
- Socaw 4: Respect another person's opinion
- Socaw 8: Cooperate with peers or siblings

#### Limitations

The sample was made up of students classified as emotionally disturbed, and therefore these findings cannot be generalized to other classified students. Furthermore, the sample did not investigate potential differential item functioning based on student characteristics, for example, classification, gender, age, and grade level. The study did not address the potential for items that load on multiple competence or blended items. Like the study of the factor structure of the Big Five personality traits [57], social and emotional learning is likely to have a complex factor structure in which items cross-load on multiple competencies. Future research is needed in order to address these limitations.

#### 5. Conclusions

Students classified as emotionally disturbed struggle to develop their self- and social awareness, which is a barrier to their success both in and outside the school setting. Addressing their challenges in these areas is critical to the program for the development of other critical competencies (i.e., self-management, goal-directed behavior, responsible decision making, etc.). The development and implementation of effective interventions for students classified as emotionally disturbed requires measures that have been validated with this population. This study provides evidence that the DESSA's self- and social awareness subscale can provide a psychometrically sound assessment of self- and social awareness among students classified as emotionally disturbed. The DESSA's self- and social awareness subscales provide a strength-based approach to the assessment of this group's needs that can be aligned to interventions delivered at various tiers (i.e., tier 1, universal programs; tier 2, targeted groups; or tier 3, individual intensive interventions) [19].

**Author Contributions:** Conceptualization, G.L.F., D.A.; methodology, G.L.F., D.A.; formal analysis, G.L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Fordham University (protocol code 1277 and 23 June 2020).

**Informed Consent Statement:** The study was conducted in established or commonly accepted educational settings, that specifically involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction [45CFR46.101(b)(4)].

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was obtained from Urban Assembly and are available David Adams (dadams@urbanassembly.org) with the permission of Urban Assembly (https://urbanassembly.org/, accessed on 17 January 2022).

Conflicts of Interest: The authors declare no conflict of interest.

## References

- Milsom, A.; Glanville, J.L. Factors Mediating the Relationship Between Social Skills and Academic Grades in a Sample of Students Diagnosed with Learning Disabilities or Emotional Disturbance. *Remedial Spec. Educ.* 2010, *31*, 241–251. [CrossRef]
- Adams, D. The application of social-emotional learning principles to a special education environment. *KEDI J. Educ. Policy* 2013, 10, 103–118.
- Espelage, D.L.; Rose, C.; Polanin, J. Social-Emotional Learning Program to Reduce Bullying, Fighting, and Victimization among Middle School Students with Disabilities. *Remedial Spec. Educ.* 2015, 36, 299–311. [CrossRef]
- Espelage, D.L.; Rose, C.; Polanin, J. Social-Emotional Learning Program to Promote Prosocial and Academic Skills among Middle School Students with Disabilities. *Remedial Spec. Educ.* 2016, 37, 323–332. [CrossRef]
- Denham, S. Keeping SEL Developmental: The Importance of a Developmental Lens for Fostering and Assessing SEL Competencies 2018; Social-Emotional Competence Assessments Work Group: Chicago, IL, USA, 2018; pp. 1–13.
- Silvia, P.J.; Duval, T.S. Objective Self-Awareness Theory: Recent Progress and Enduring Problems. Pers. Soc. Psychol. Rev. 2001, 5, 230–241. [CrossRef]
- Sheldon, K.M.; Johnson, J.T. Forms of Social Awareness: Their Frequency and Correlates. *Pers. Soc. Psychol. Bull.* 1993, 19, 320–330. [CrossRef]

- 8. Ross, K.M.; Tolan, P. Social and Emotional Learning in Adolescence: Testing the CASEL Model in a Normative Sample. *J. Early Adolesc.* **2017**, *38*, 1170–1199. [CrossRef]
- Denham, S.A.; Bassett, H.H.; Zinsser, K.; Wyatt, T.M. How Preschoolers' Social-Emotional Learning Predicts Their Early School Success: Developing Theory-Promoting, Competency-Based Assessments. *Infant Child Dev.* 2014, 23, 426–454. [CrossRef]
- Falon, S.L.; Kangas, M.; Crane, M.F. The coping insights involved in strengthening resilience: The Self-Reflection and Coping Insight Framework. *Anxiety Stress Coping* 2021, 34, 734–750. [CrossRef]
- 11. Payton, J.; Roger, W.; Joseph, D. *The Positive Impact of Social and Emotional Learning for Kindergarten to Eighth-Grade Students: Findings from Three Scientific Reviews*; Technical Report; Collaborative for Academic, Social, and Emotional Learning: Chicago, IL, USA, 2008.
- Schonfeld, D.J.; Adams, R.E.; Fredstrom, B.K.; Weissberg, R.P.; Gilman, R.; Voyce, C.; Tomlin, R.; Speese-Linehan, D. Clusterrandomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *Sch. Psychol. Q.* 2015, 30, 406–420. [CrossRef]
- 13. NCSL. Social and Emotional Learning: State Legislation. 2021. Available online: https://www.ncsl.org/research/education/ social-emotional-learning.aspx (accessed on 1 January 2021).
- 14. Daley, S.G.; McCarthy, M.F. Students With Disabilities in Social and Emotional Learning Interventions: A Systematic Review. *Remedial Spec. Educ.* **2020**, *42*, 384–397. [CrossRef]
- Hagarty, I.; Morgan, G. Social-emotional learning for children with learning disabilities: A systematic review. *Educ. Psychol. Pract.* 2020, *36*, 208–222. [CrossRef]
- Ura, S.K.; Castro-Olivo, S.M.; D'Abreu, A. Outcome Measurement of School-Based SEL Intervention Follow-Up Studies. Assess. Eff. Interv. 2019, 46, 76–81. [CrossRef]
- 17. LeBuffe, P.A.; Shapiro, V.B.; Naglieri, J.A. *DESSA Devereux Student Strengths Assessment K-8th Grade*; Apperson SEL+: Charlotte, NC, USA, 2014.
- Nickerson, A.B.; Fishman, C. Convergent and divergent validity of the Devereux Student Strengths Assessment. *Sch. Psychol. Q.* 2009, 24, 48–59. [CrossRef]
- 19. Haggerty, K.; Elgin, J.; Woolley, A. *Social-Emotional Learning Assessment Measures for Middle School Youth*; Rakees Foundation: Seattle, WA, USA, 2011.
- 20. LeBuffe, P.A.; Shapiro, V.B.; Robitaille, J.L. The Devereux Student Strengths Assessment (DESSA) comprehensive system: Screening, assessing, planning, and monitoring. *J. Appl. Dev. Psychol.* **2018**, *55*, 62–70. [CrossRef]
- 21. Yang, F.M.; Kao, S.T. Item response theory for measurement validity. Shanghai Arch. Psychiatry 2014, 26, 171–177. [CrossRef]
- 22. Nicewander, A. Conditional Reliability Coefficients for Test Scores. *Psychol. Methods* 2017, 23, 351–362. [CrossRef]
- 23. Embretson, S.E.; Reise, P.S. Chapter 5. Polytomous IRT Models. In *Item Response Theory*; Psychology Press: Mahwah, NJ, USA, 2000.
- 24. Susan, E.E.; Steven, P.R. *Item Response Theory for Psychologists*; Multivariate Applications Book Series; Psychology Press: Mahwah, NJ, USA, 2000.
- Zanon, C.; Hutz, C.; Yoo, H.; Hambleton, R.K. An application of item response theory to psychological test development. *Psicologia Reflexão e Crítica* 2016, 29, 18. [CrossRef]
- Reise, S.P.; Ainsworth, A.T.; Haviland, M.G. Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. *Curr. Dir. Psychol. Sci.* 2005, 14, 95–101. [CrossRef]
- 27. Samejima, F. Chapter 4: The General Graded Response Model; Ostini, R., Nering, M.L., Eds.; Routledge: New York, NY, USA, 2010.
- 28. Reise, S.P.; Waller, N.G. Item Response Theory and Clinical Measurement. Annu. Rev. Clin. Psychol. 2009, 5, 27–48. [CrossRef]
- 29. Toland, M.D. Practical Guide to Conducting an Item Response Theory Analysis. J. Early Adolesc. 2013, 34, 120–151. [CrossRef]
- Bean, G.J.; Bowen, N.K. Item Response Theory and Confirmatory Factor Analysis: Complementary Approaches for Scale Development. J. Evid.-Based Soc. Work 2021, 18, 597–618.
- 31. Edelen, M.O.; Reeve, B.B. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual. Life Res.* 2007, *16*, 5–18. [CrossRef] [PubMed]
- National Center on Birth Defects and Developmental Disabilities. Autism and Developmental Disabilities Monitoring Network— 2012. 2012; pp. 1–45. Available online: http://www.cdc.gov/ncbddd/autism/documents/ADDM-2012-Community-Report.pdf (accessed on 13 January 2021).
- 33. IBM Corp. IBM SPSS Statistics for Windows; Verizon 25.0; IBM Corp: Armonk, NY, USA, 2017.
- 34. Lorenzo-Seva, U.; Ferrando, V.-C. FACTOR 10.10.03. 2020. Available online: http://psico.fcep.urv.es/utilitats/factor/Download. html (accessed on 2 October 2020).
- R Core Team. R: A Language and Environment for Statistical Computing Computing. 2016. Available online: Https://www.R-project.org/ (accessed on 1 March 2020).
- 36. Desjardins, C.D.; Bulut, O. *Handbook of Educational Measurement and Psychometrics Using R*; Taylor & Francis Group: Abingdon, UK, 2018.
- Forero, C.G.; Maydeu-Olivares, A. Estimation of IRT graded response models: Limited versus full information methods. *Psychol. Methods* 2009, 14, 275–299. [CrossRef] [PubMed]
- Wang, C.; Su, S.; Weiss, D.J. Robustness of Parameter Estimation to Assumptions of Normality in the Multidimensional Graded Response Model. *Multivar. Behav. Res.* 2018, 53, 403–418. [CrossRef]

- Immekus, J.C.; Snyder, K.E.; Ralston, P.A. Multidimensional Item Response Theory for Factor Structure Assessment in Educational Psychology Research. *Front. Educ.* 2019, 4, 45. [CrossRef]
- 40. Nguyen, T.H.; Han, H.-R.; Kim, M.T.; Chan, K.S. An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *Patient Patient-Cent. Outcomes Res.* 2014, 7, 23–35. [CrossRef]
- Lubbe, D.; Schuster, C. A Graded Response Model Framework for Questionnaires With Uniform Response Formats. *Appl. Psychol. Meas.* 2018, 43, 290–302. [CrossRef]
- 42. Baker, F.B.; Kim, S.H. Chapter 1: The item characteristic curve. In *The Basic of Item Response Theory in R*; Springer International Publishing: New York, NY, USA, 2017.
- 43. Mokken, R.J. A Theory and Procedure of Scale Analysis: With Application in Political Research; The Hague: Mouton: Berlin, Germany, 1971; p. 353.
- 44. Ark, L. Mokken scale analysis in R. J. Stat. Softw. 2007, 20, 1–19. [CrossRef]
- 45. Timmerman, M.E.; Lorenzo-Seva, U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol. Methods* **2011**, *16*, 209–220. [CrossRef]
- 46. Ruscio, J.; Roche, B. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychol. Assess.* **2012**, *24*, 282–292. [CrossRef]
- 47. Lim, S.; Jahng, S. Determining the number of factors using parallel analysis and its recent variants. *Psychol. Methods* **2019**, *24*, 452–467. [CrossRef]
- Ferrando, P.J.; Lorenzo-Seva, U. Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory factor score estimates in exploratory item factor analysis. *Educ. Psychol. Meas.* 2018, 78, 762–780. [CrossRef]
- 49. Shi, D.; Lee, T.; Maydeu-Olivares, A. Understanding the Model Size Effect on SEM Fit Indices. *Educ. Psychol. Meas.* 2018, 79, 310–334. [CrossRef] [PubMed]
- 50. Chen, W.H.; Thissen, D. Local dependence indices for item pairs using item response theory. J. Educ. Behav. Stat. 1997, 22, 265–289. [CrossRef]
- Edwards, M.C.; Houts, C.R.; Cai, L. A diagnostic procedure to detect departures from local independence in item response theory models. *Psychol. Methods* 2018, 23, 138–149. [CrossRef] [PubMed]
- 52. Maydeu-Olivares, A. Goodness-of-Fit Assessment of Item Response Theory Models. *Meas. Interdiscip. Res. Perspect.* 2013, 11, 71–101. [CrossRef]
- 53. Hu, L.T.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* **1999**, *6*, 1–55. [CrossRef]
- 54. Xu, J.; Paek, I.; Xia, Y. Investigating the Behaviors of M2 and RMSEA2 in Fitting a Unidimensional Model to Multidimensional Data. *Appl. Psychol. Meas.* **2017**, *41*, 632–644. [CrossRef]
- 55. Baker, F.B. *The Basics of Item Response Theory*; ERIC Clearinghouse on Assessment and Evaluation, University of Marylandl: College Park, MD, USA, 2001.
- Andersson, B.; Xin, T. Large Sample Confidence Intervals for Item Response Theory Reliability Coefficients. *Educ. Psychol. Meas.* 2017, 78, 32–45. [CrossRef]
- 57. Schmidt, F.; DeGeest, D.S. A Rigorous Test of the Fit of the Circumplex Model to Big Five Personality Data: Theoretical and Methodological Issues and Two Large Sample Empirical Tests. *Multivar. Behav. Res.* **2015**, *50*, 350–364. [CrossRef]