Taylor & Francis
Taylor & Francis Group

Journal of Intellectual & Developmental Disability

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjid20

An evaluation of the German teacher version of the Developmental Behaviour Checklist in children and adolescents with intellectual disability

Carmen L. A. Zurbriggen & Christoph M. Müller

To cite this article: Carmen L. A. Zurbriggen & Christoph M. Müller (2022): An evaluation of the German teacher version of the Developmental Behaviour Checklist in children and adolescents with intellectual disability, Journal of Intellectual & Developmental Disability, DOI: 10.3109/13668250.2022.2044269

To link to this article: https://doi.org/10.3109/13668250.2022.2044269

JOURNAL OF INTELLECTUAL & DEVELOPMENTAL DISABILITY
https://doi.org/10.3109/13668250.2022.2044269

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

An evaluation of the German teacher version of the Developmental Behaviour Checklist in children and adolescents with intellectual disability

Carmen L. A. Zurbriggen [a] and Christoph M. Müller [b]

[a]Faculty of Humanities, Education and Social Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg; [b]Department of Special Education, University of Fribourg, Fribourg, Switzerland

ABSTRACT

Background: The Developmental Behaviour Checklist (DBC) is an established, internationally used questionnaire for assessing behavioural and emotional problems among young people with developmental or intellectual disabilities (ID). The present study aims to evaluate the psychometric properties of its German teacher version (DBC-T).

Method: The German DBC-T was administered to 397 school staff members who reported twice on 1177 children and adolescents with ID over a period of 7–9 months. Data were analysed within an exploratory structural equation modelling framework.

Results: Our results supported the five-factor structure of the DBC-T and found good reliability for all scales. Analyses on the relationship of DBC-T scores with students' age, gender, and adaptive behaviour provided further evidence for the validity of the DBC-T.

Conclusions: Our study endorses the notion that the German DBC-T is an important instrument both for research and practice. Limitations and further directions are discussed.

KEYWORDS

Intellectual disability; assessment; behaviour; Developmental Behaviour Checklist; evaluation

In addition to exhibiting difficulties in intellectual and adaptive functioning, children and adolescents with intellectual disabilities (ID) often show increased rates of behavioural and emotional problems (American Association on Intellectual and Developmental Disabilities, 2021; Dekker et al., 2002c). According to Einfeld and Tonge (1995), an individual's behaviours and emotions are considered disordered when there is a qualitative or quantitative deviance that cannot be explained by developmental delay alone, that causes distress to the person or to caregivers, and that is a significant additional impairment to adaptive functioning. A broad spectrum of behavioural and emotional difficulties can be seen in ID, such as disruptive, antisocial, and self-absorbed behaviours; communication disturbances; anxiety; and difficulties relating socially (Einfeld & Tonge, 1995). Dekker et al. (2002c) found a three- to four-fold elevated risk for overall problem behaviour in children and adolescents with ID compared to those without ID.

Children and adolescents with ID who exhibit challenging behaviours are at risk for injury, controversial medication and restraint, and social isolation; in addition, they often rely on residential service provision (Emerson & Einfeld, 2011). These difficulties can also be very stressful for parents, caretakers, and teachers (e.g., Amstad & Müller, 2020; Hastings, 2002). Thus, it is important to be able to identify the degree and topography of emotional and behavioural problems, using valid and reliable instruments, so that effective intervention can occur. For instance, based on a specific individual profile of strengths and difficulties, school and clinical staff may decide on certain emotional and behavioural problems that intervention should primarily focus on.

Several instruments to assess behavioural and emotional problems in children and adolescents with ID currently exist. According to a recent systematic review (Buckley et al., 2020), the Developmental Behaviour Checklist (DBC) published by Einfeld and Tonge (1995, 2002) is internationally among the two most commonly the most commonly used symptom rating scales. The DBC exists in various forms for the assessment of both adults (DBC-A) and children or adolescents with ID. For children and adolescents, the questionnaire can be filled out by parents and carers (DBC-P) or by teachers (DBC-T), with a high overlap in items between the two versions.

CONTACT Carmen L. A. Zurbriggen carmen.zurbriggen@uni.lu

Supplemental data for this article can be accessed at https://doi.org/10.3109/13668250.2022.2044269.

The present study focuses on the DBC-T, which provides important information on the degree to which emotional and behavioural problems of children and adolescents with ID are perceived by school staff. The DBC-T has been successfully evaluated for use in several countries (e.g., Dekker et al., 2002a, 2002b). Einfeld et al. (2007) translated the DBC-T, together with the DBC-P and DBC-A, for use in German-speaking countries. The authors evaluated the German DBC-P and found it possessed high psychometric quality (Steinhausen & Winkler Metzke, 2005). However, to date, little knowledge exists on the psychometric properties of the DBC-T, particularly for the German version. Hence, the present paper aims to shed additional light on this instrument's validity and reliability, using data from a longitudinal study on children and adolescents with ID.

Psychometric properties of the DBC-T

The DBC was developed by Einfeld and Tonge (1995), adapting the model of the Child Behaviour Checklist (CBCL; Achenbach, 1991), though specifically designed to assess behavioural and emotional problems in people with developmental and intellectual disabilities. The instrument exists in different versions, which all are derived from or based on the DBC-P (for an overview, see Einfeld & Tonge, n.d.). The DBC-P comprises 96 items, each corresponding to a behavioural description that is rated by parents or caregivers, who are asked to make their assessment using a retrospective period of the prior six months. As with the DBC-P, the DBC-T is intended for children and adolescents aged 4–18 (Einfeld et al., 1999; Einfeld & Tonge, 2002). Both versions are very similar and primarily differ in terms of item number: Three DBC-P items that assess sleeping behaviour are not included in the DBC-T, and one school-context specific item was added to the DBC-T, resulting in 94 items for the DBC-T. More recently, a revised online version of the DBC was published (DBC-2; Gray et al., 2018).

The original six-factor structure of the DBC (Einfeld & Tonge, 1992, 1995) was re-examined based on a large Australian-Dutch sample in a study by Dekker et al. (2002a). For the DBC-P, results revealed five clearly interpretable and internally consistent subscales: disruptive/antisocial, self-absorbed, communication disorders, social relating, and anxiety. However, for the DBC-T, the principle components analysis failed to identify a separate and meaningful factor for anxiety. Furthermore, the internal consistency of the anxiety subscale was marginally less than satisfactory. Given the broad similarities between both DBC versions and in order to facilitate comparisons, the five-factor structure was also retained for the teacher version (Dekker et al., 2002a). It should be noted that some items were excluded from the principal component analyses in both versions of the DBC due to low factor loadings, low observed frequencies in certain categories, or because they pointed to behaviours symptomatic of psychotic illness. These items were retained for the calculation of the Total Behaviour Problem Score (TBPS), due to their relevance in assessing mental disorders in the context of ID. Nevertheless, based on their findings, the authors suggested to replicate the factor structure and to consider the development of a shortened version of the DBC.

Fewer findings are available on the validity of the DBC-T than the DBC-P. Therefore, in the following, we also point to results on the latter. Convergent validity was shown by high correlations between the TBPS of the DBC-T and the CBCL (Dekker et al., 2002b), and between the TBPS of the DBC-P and the Strengths and Difficulties Questionnaire's (Goodman, 2001) total difficulties score (Rice et al., 2018). Furthermore, Dekker et al. (2002b) provided evidence for discriminant validity with adaptive behaviour, as measured by the Vineland Adaptive Behaviour Scales (VABS; Sparrow et al., 1984). The three DBC-T scales disruptive/antisocial, anxiety, and social relating showed small negative correlations with the three VABS domains communication, socialisation, and daily living skills, as well as with total adaptive functioning, while the TBPS and the scales self-absorbed and communication disturbances were moderately correlated with adaptive behaviour. Overall, the results suggested that the DBC-T and the VABS cover related but different concepts (Dekker et al., 2002b), thus providing support for discriminant validity of the DBC-T.

In terms of construct validity, group differences can be interpreted as known group validity. One frequently investigated variable in terms of group differences in gender (see, e.g., review by McKenzie et al., 2016). Using the DBC-T, Molteno et al. (2001) reported higher levels of emotional and behavioural problems in boys than girls, as expected.

For young people with ID, certain studies suggest a slight decrease in problem behaviour as they progress from childhood to adolescence (e.g., de Ruiter et al., 2007; Einfeld et al., 2006), thus age or age-related development can serve as further indicator for the validation of the DBC. de Ruiter et al. (2008) found a decrease in parent-reported problem behaviours over a five-year period, except for the social relating scale. No significant changes were found for teacher ratings for all scales. In turn, the findings on the DBC-T by Molteno et al. (2001) showed that adolescents had fewer behaviour

problems compared to younger children. Studies that considered a shorter time frame reported either relative persistence (i.e., one-year stability; Dekker et al., 2002a), or a slight reduction in problem behaviour after an average period of 14 months (Chandler et al., 2015).

The German version of the DBC-T

The German version of the DBC is called *Verhaltensfragenbogen bei Entwicklungsstörungen* (VFE; Einfeld et al., 2007). Information in the German manual concerning the psychometric properties of the DBC-T is based on the Australian samples used to develop or evaluate the English DBC-T and DBC-P (Einfeld & Tonge, 1995, 2002). In addition, the results of a study by Steinhausen and Winkler Metzke (2005) on the evaluation of the German DBC-P are reported. This study was based on a sample of 721 children and adolescents with ID in Germany and included various psychometric analyses as well as the construction of norm tables according to different severity levels of ID. The results replicated for the German DBC-P version the five-factor structure of the DBC as revised by Dekker et al. (2002a, 2002b), and overall found the internal consistency of the scales was adequate to good ($|\alpha| = 0.69–0.90$), except for the anxiety scale with Cronbach's α of 0.58 (Steinhausen & Winkler Metzke, 2005). Retest reliability after a period of 1.8 years was high for all five subscales.

Since the release of the VFE manual in 2007, the adult form of the German DBC (DBC-A or VFE-ER, respectively) has been evaluated and standardised in a study by Steinhausen and Winkler Metzke (2011). According to our research, no study has yet been published evaluating the German DBC-T. Although the differences in scale construction and item wording between the DBC-P and the DBC-T are minor, findings comparing the English DBC-P and DBC-T indicated that parent and teacher ratings are only partially consistent. In a further evaluation of the Dutch DBC (Dekker et al., 2002b), the agreement between parent and teacher ratings was shown to be low to moderate, with lowest agreement for the anxiety scale ($r = 0.27$) and highest for the self-absorbed scale ($r = 0.57$). Chandler et al. (2005) reported rather large discrepancies between parental and teacher reports of overall problem severity as measured by the English version of the DBC: Only 64% of children whose scores were above the cut-off on the DBC-P were also above the cut-off on the DBC-T. The low to moderate agreement between DBC-P and DBC-T ratings is in line with other studies investigating multi-informant reports on children's or adolescents' behavioural and emotional problems (for an overview, see: De Los Reyes et al., 2015; Rescorla et al., 2014). Thus, parent ratings using the DBC-P and teacher ratings using the DBC-T may provide different information, which can result in differing findings (e.g., de Ruiter et al., 2008), and, in turn, may lead to different conclusions. While inconsistencies in ratings between parents and teachers can provide important information both for research and practice, it is crucial, first, to examine the German version of the DBC-T, in order to obtain more precise information on its validity and reliability.

Aims of the present study

This study aimed to evaluate the German version of the DBC-T (Einfeld et al., 2007). For this purpose, we used data from a longitudinal study conducted in special needs schools for students with ID, where school staff had reported on students' emotional and behavioural problems using the German DBC-T. First, we examined the dimensionality and factor structure of the instrument. Second, we assessed the reliability (i.e., internal consistency) of the subscales. Third, we evaluated different aspects of validity by investigating differences across gender and age (i.e., known group validity as one aspect of construct validity) and by predicting DBC-T scores by students' adaptive behaviour levels (i.e., discriminant validity).

Method

Sample and procedure

The data originated from the Swiss longitudinal study "Competent with Peers – ComPeers" (funded by the Swiss National Science Foundation under Grant SNF-172773), which was approved in terms of scientific and ethical conduct by the institutional research commission of the relevant university department. The first measurement occasion (T1) took place between 1 and 2 month(s) after the start of a school year, the second measurement occasion (T2) 7–9 months later, near the end of the same school year. The sample was drawn from special needs schools who serve students with ID in the German-speaking part of Switzerland. These schools can only be attended by students who have a clinical diagnosis of ID. Diagnoses are usually based on ICD-10 criteria, including an assessment of intelligence (IQ < 70) and a clinical estimation of adaptive behaviour levels. Parents were informed in writing about the study. The letter emphasised that all information would be assessed anonymously (i.e., researchers did not have access to the names of the students, parents, or staff filling out the questionnaires) and that

no information on medical diagnoses would be collected. Parents and teaching staff had the option to decline participation.

The sample consisted of 1177 children and adolescents (65.9% girls) with a mean age at T1 of 11.26 years (SD_{age} = 2.76, min = 4.17, max = 19.08). This corresponds to about 96% of the student body from the 16 special needs schools included. In total, 397 staff members (M_{age} = 46.26 years, SD_{age} = 12.53, 86.6% female) took part in the study and reported on students who attended the classroom in which they worked. Each staff member completed questionnaires for an average of 2.69 children or adolescents (*SD* = 1.62, range = 1–8). The majority of staff members (61.5%) held a diploma in special needs education; others were regular teachers, therapists, social workers, or long-term trainees (for more information, see Müller et al., 2021).

Measures

Emotional and behavioural problems

Emotional and behavioural problems in children and adolescents were assessed using the teacher version of the VFE (Einfeld et al., 2007), namely the German version of the DBC-T (Einfeld et al., 1999; Einfeld & Tonge, 1995, 2002). For each of the 94 items, school staff members reported on a student's behaviour over the last two months on a 3-point rating scale (0 = *not true as far as you know*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*). A two-month reporting period for observed behaviour was used (instead of six month) to guarantee valid ratings by staff who had only known students since the beginning of the school year. The last item (Item 94) corresponds to an overall rating of whether the child has problems with feelings or behaviour, in addition to problems with development. This item only is rated based on different category labels (0 = *no*, 1 = *yes, but minor*, 2 = *yes, major*).

Adaptive behaviour

Adaptive behaviour was assessed using a German version of the Adaptive Behaviour Assessment System 3 (ABAS-3; Bienstein et al., 2017). The ABAS-3 consists of 174 items assessing the conceptual, social, and practical skills required to function in daily life. We used the teacher form intended for children or adolescents between 5 and 21 years old. School staff were asked to score each item on a 4-point rating scale (0 = *is not able*, 1 = *never or almost never when needed*, 2 = *sometimes when needed*, 3 = *always or almost always when needed*).

In the present study, we employed the percentile ranks of the General Adaptive Composite (GAC) score as well as of the three adaptive domains conceptual, social, and practical skill areas, based on age-related norms of the original English ABAS-3 (Harrison & Oakland, 2015), as there are no norms available for the German version yet.

Analyses

All analyses were performed in M*plus* Version 8.6 (Muthén & Muthén, 1998–2017). Parameters were estimated with the robust weighted least squares mean- and variance-adjusted (WLSMV) estimator. The WLSMV is specifically designed for categorical data (see recommendation by Li, 2015; Nussbeck et al., 2006), in which the normality assumption is typically violated, which has to be assumed for the behavioural data measured with the DBC.

First, we performed a series of confirmatory factor analyses (CFA) with ordered categorical factor indicators, in order to examine the dimensional structure of the German DBC-T. More specifically, we tested the following models: CFA with one general factor (Model 1); CFA with five factors in line with the given subscales of the DBC-T (Model 2); CFA with five factors and the cross-loadings indicated in the manual of the German DBC-T (Model 3); CFA with six factors including the nine additional items, which are considered for the overall score but not included in the five original subscales (Model 4); and CFA with six factors and cross-loadings as indicated in the manual (Model 5).

As per CFA specifications, items are allowed to only load on the hypothesised factor, while cross-loadings are constrained to zero. These assumptions, however, are often overly restrictive for multifactor rating instruments routinely used in applied research, so that their structure cannot be represented adequately within a CFA approach. The associated failing of meeting the goodness-of-fit criteria has led to a tendency toward considerable model modifications to find a well-fitting model (Asparouhov & Muthén, 2009). Common consequences of the strict requirement in CFA include distorted factors with inflated factor loadings (Marsh et al., 2009). In the light of these issues, and since previous research suggested several cross-loadings for the DBC-T, we additionally applied an exploratory factor structural equation modelling (ESEM; Asparouhov & Muthén, 2009) approach. ESEM includes the advantages of both EFA and CFA within a latent variable modelling framework, giving access to the advanced statistical applications of CFA and structural equation modelling (Marsh et al., 2014; Morin et al., 2018). Hence, in addition to the CFA models, we specified an ESEM with five target factors (Model 6) in line with the

given subscales of the German DBC-T, and an ESEM with a sixth target factor, including the nine additional items (hereafter described as others factor; Model 7). The ordered categorical indicators loaded on their specific (main) factors, while cross-loadings with other factors were targeted to be as close to zero as possible (but not forced to zero) with the oblique target rotation procedure (Browne, 2001; for an example, see Tóth-Király et al., 2017).

All CFA and ESEM models were first fitted on data from T1 and, in a second phase of the analyses, on data from T2 to test replicability. To assess the fit of the CFA and ESEM models, we relied on the following sample size-independent goodness-of-fit indices: the comparative fit index (CFI), the Tucker–Lewis index (TLI), the standardised root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). As guidelines, TLI and CFI values greater than 0.90 and 0.95 are considered adequate and excellent model fit, respectively. For the SRMR, values below 0.08 are generally recommended. RMSEA values less than 0.06 reflect close fit to the data (Hu & Bentler, 1999; Marsh et al., 2004). For the sake of completeness, we report the chi-square test (χ^2) of exact fit as well, despite its known sensitivity to large sample size, model complexity and non-normal data (Yuan & Bentler, 2004). Since model assessment should not only be based on goodness-of-fit statistics, we also inspected parameter estimates, in particular the factor loadings (Morin et al., 2018).

The internal consistency of the German DBC-T subscales was evaluated by calculating McDonald's ω (McDonald, 2013) within the latent modelling framework. Compared to other indices, such as Cronbach's α, ω provides a more accurate approximation of internal consistency (Dunn et al., 2014; Sijtsma, 2009). In general, reliability coefficients of 0.80 or greater indicate good internal consistency, and values around 0.70 are considered adequate (Kline, 2015).

To examine different aspects of validity, we applied a multiple indicator multiple causes (MIMIC) model within the ESEM approach (Marsh et al., 2013). Basically, the MIMIC approach corresponds to a multivariate regression model in which latent variables are regressed on predictors (Morin et al., 2018). In the ESEM MIMIC analyses used in our study, the categorical variable gender (female vs. male), as well as the continuous variables age and adaptive behaviour (GAC and the three domains conceptual, social, practical skills) were added separately to the ESEM model at T1 and T2. In a more exploratory manner, we additionally investigated the predictive value of the overall rating of behavioural and emotional problems (assessed with item 94 of the DBC-T).

Results

Factorial structure of the DBC-T

Table 1 reports the goodness-of-fit statistics for the CFA and ESEM models. In general, the CFA models had unsatisfactory model fit at T1, as apparent by the fit indices (Models 1.1–1.5). Allowing the cross-loadings as indicated by the manual of the German DBC (Models 1.3 and 1.5) still resulted in unsatisfactory fit indices. In turn, both ESEM models (Models 1.6 and 1.7) showed an appropriate fit to the data. At T2, the results followed the same pattern: While the CFA models did not fit the data well, the ESEM models reached acceptable to good fit. Overall, the ESEM solutions with six factors resulted in the best fitting model.

Table 1. Summary of goodness-of-fit statistics of the CFA and ESEM models.

Model	χ^2	df	*p*	CFI	TLI	SRMR	RMSEA [90% CI]
T1							
(1.1) CFA 1 factor	22,037.74	4277	<.001	0.674	0.667	0.148	0.062 [0.061, 0.062]
(1.2) CFA 5 factors[a]	14,259.44	3310	<.001	0.779	0.773	0.118	0.055 [0.054, 0.056]
(1.3) CFA 5 factors with cross-loadings[a,b]	13,515.73	3306	<.001	0.794	0.788	0.115	0.053 [0.052, 0.054]
(1.4) CFA 6 factors	16,290.03	4170	<.001	0.771	0.765	0.130	0.052 [0.051, 0.052]
(1.5) CFA 6 factors with cross-loadings[b]	19,323.53	4259	<.001	0.725	0.718	0.137	0.057 [0.056, 0.057]
(1.6) ESEM 5 factors[a]	6089.51	2998	<.001	0.938	0.929	0.055	0.031 [0.030, 0.032]
(1.7) ESEM 6 factors	6265.19	3649	<.001	0.950	0.942	0.054	0.026 [0.025, 0.027]
T2							
(2.1) CFA 1 factor	20,769.97	4185	<.001	0.669	0.661	0.146	0.061 [0.060, 0.062]
(2.2) CFA 5 factors[a]	13,781.21	3310	<.001	0.780	0.774	0.121	0.055 [0.054, 0.055]
(2.3) CFA 5 factors with cross-loadings[a,b]	12,405.09	3306-	<.001	0.657	0.647	0.082	0.051 [0.050, 0.052]
(2.4) CFA 6 factors	15,388.59	4170	<.001	0.776	0.770	0.125	0.050 [0.049, 0.051]
(2.5) CFA 6 factors with cross-loadings[b]	14,554.92	4166	<.001	0.642	0.632	0.080	0.048 [0.048, 0.049]
(2.6) ESEM 5 factors[a]	5547.14	2998	<.001	0.946	0.939	0.055	0.028 [0.027, 0.029]
(2.7) ESEM 6 factors	5737.50	3649	<.001	0.958	0.952	0.052	0.023 [0.022, 0.024]

Notes: CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling; CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardised root mean square residual; RMSEA = root mean square error of approximation, 90% CI = 90% confidence interval. [a]additional items ("others") not included; [b]with cross-loadings as indicated in the manual for the German DBC-T.

In addition to the goodness-of-fit statistics, we inspected parameter estimates, in particular the factor loadings (Morin et al., 2018). The results for the CFA models with five or six factors showed generally moderate to substantial factor loadings, with a few exceptions of weak loadings ($\lambda < 0.30$), at both T1 and T2 (e.g., Model 1.2, disruptive/antisocial factor: $|\lambda| = 0.209$–0.867, $M = 0.622$). The ESEM results yielded generally well-defined factors for the DBC-T, although again with a wide range of magnitude in targeted factor loadings (e.g., Model 1.6, disruptive/antisocial factor: $|\lambda| = 0.192$–0.889, $M = 0.648$). As per ESEM specification, the items were also allowed to load on the other factors, thus cross-loadings were present. In the ESEM with five factors, for instance, four cross-loadings were significant and substantial ($\lambda \geq 0.30$) for the disruptive/antisocial scale. The ESEM model with six factors revealed several nonsignificant, weak factor loadings for the sixth targeted factor at T1 and T2 (e.g., Model 1.7, others factor: $|\lambda| = 0.009$–0.356, $M = 0.188$), suggesting the additional items do not represent one common factor. Table S1 in the Supplementary Material displays the standardised factor loadings and standard errors in the ESEM model with six factors at T1. The corresponding items loaded substantially on the targeted factor, with one to three exceptions per factor. As expected by specification in the ESEM model, cross-loadings were present for all factors. For the others factor, several cross-loadings were quite substantial, indicating again that the indicators tap into a construct different from the targeted factor. This conclusion was supported by the very small, nonsignificant correlations with the five original subscales (see Table 2), except for the self-absorbed scale, which was negatively correlated with the others factor at T1, even though this effect was small in magnitude (−0.13, $p < .001$).

Regarding the five original scales, the intercorrelations in the ESEM model were all significant at both measurement occasions. As reported in Table 2, they varied between low to moderate, with the lowest correlations between social relating and disruptive/antisocial at both T1 and T2 (T1: 0.06, $p < .05$ / T2: 0.10, $p < .001$), and the highest correlations were between communication disturbance and self-absorbed (T1: 0.32, $p < .001$ / T2: 0.35, $p < .001$), and at T2 between anxiety and disruptive/antisocial as well (T2: 0.35, $p < .001$). Thus, the results related to the ESEM model with six factors suggested that the five original scales of the DBC-T are moderately related, which can be interpreted as moderate discriminant validity of the scales. The additional items (included in the others scale), however, might constitute a separate construct and can be considered as items with independent information.

In sum, based on all our results regarding factorial structure, we concluded that the DBC-T is best represented by the ESEM model with five factors, thus without the additional items.

Reliability of the DBC-T scales

The reliabilities of the German DBC-T subscales are available in the main diagonals (in parentheses) of Table 2. The McDonalds ω coefficients ranged from $\omega = 0.85$ (95% confidence interval (CI) = [0.83, 0.86]) to $\omega = 0.96$ (95% CI = [0.96, 0.97]) at T1, and from $\omega = 0.82$ (95% CI = [0.81, 0.84]) to $\omega = 0.96$ (95% CI = [0.96, 0.97]) at T2, indicating good internal consistency of all subscales at both measurement occasions. Even the scale with the additional items (others scale) reached adequate internal consistency, although with the lowest internal consistency of $\omega = 0.75$ (95% CI = [0.72, 0.78]) at T1, and $\omega = 0.77$ (95% CI = [0.74, 0.80]) at T2.

Validity of the DBC-T

In light of the results regarding the factorial structure of the DBC-T, we applied the MIMIC analysis on the ESEM model with five factors (i.e., without the additional items included in the others scale). The standardised regression coefficients of the predictors on the latent factors in the ESEM MIMIC model are presented in Table 3. The standardised estimates can be interpreted as effects sizes, akin to Cohen's *d* (Brown, 2015).

Gender had a small to medium effect on the five factors at T1. Boys scored higher than girls, except regarding anxiety (−0.279, $p < .001$). At T2, the effect of gender was somewhat less pronounced, and for the

Table 2. Correlations in the ESEM models with six factors and reliabilities (McDonald ω) of the latent factors of the German DBC-T.

	DISR	SELF	COM	ANX	SOC	OTH
T1						
DISR	(0.95)					
SELF	0.30***	(0.96)				
COM	0.19***	0.32***	(0.89)			
ANX	0.27***	0.26***	0.21***	(0.85)		
SOC	0.06**	0.25***	0.14***	0.26***	(0.85)	
OTH	0.05	−0.13***	−0.02	0.04	−0.04	(0.75)
T2						
DISR	(0.95)					
SELF	0.24***	(0.96)				
COM	0.22***	0.35***	(0.90)			
ANX	0.35***	0.19***	0.23***	(0.82)		
SOC	0.10***	0.30***	0.21***	0.22***	(0.84)	
OTH	0.03	−0.05	−0.02	0.03	−0.05	(0.77)

Notes: DISR = discruptive/antisocial, SELF = self-absorbed, COM = communiation disturbance, ANX = anxiety, SOC = social relating, OTH = others. In parentheses (main diagonals): reliability coefficients (McDonald ω). For identification of the mean structure, the first threshold of each factor has been fixed to 0. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3. Standardised regression coefficients (and standard errors) on the latent factors in the ESEM MIMIC model.

	DISR	SELF	COM	ANX	SOC
T1					
Gender	0.219*** (0.069)	0.188** (0.077)	0.397*** (0.079)	−0.297*** (0.077)	0.335*** (0.079)
Age	−0.093** (0.032)	−0.281*** (0.031)	−0.149*** (0.037)	−0.070 (0.036)	−0.014 (0.035)
GAC	−0.155** (0.034)	−0.619*** (0.019)	−0.252*** (0.042)	−0.179*** (0.044)	−0.076* (0.038)
Conceptual skills	0.050 (0.062)	0.036 (0.052)	−0.233*** (0.068)	−0.014 (0.067)	−0.015 (0.053)
Social skills	0.005 (0.048)	−0.309*** (0.039)	0.028 (0.058)	−0.046 (0.058)	−0.432*** (0.049)
Practical skills	−0.226*** (0.065)	−0.376*** (0.052)	−0.168*** (0.074)	−0.109 (0.074)	0.158** (0.060)
T2					
Overall assessment[a] T1	0.363*** (0.029)	0.312*** (0.032)	0.289*** (0.034)	0.223*** (0.041)	0.364*** (0.031)
Gender	0.107*** (0.032)	0.060 (0.037)	0.204*** (0.034)	−0.111** (0.038)	0.105** (0.035)
Age	−0.125*** (0.032)	−0.275*** (0.031)	−0.146*** (0.037)	−0.055 (0.036)	−0.019 (0.035)
GAC	−0.233*** (0.031)	−0.747*** (0.015)	−0.177*** (0.053)	−0.362*** (0.044)	−0.156*** (0.048)
Conceptual skills	−0.029 (0.056)	0.043 (0.044)	−0.226*** (0.068)	−0.105 (0.074)	0.079 (0.061)
Social skills	−0.065 (0.047)	−0.350*** (0.033)	−0.099 (0.059)	−0.144* (0.017)	−0.604*** (0.045)
Practical skills	−0.154* (0.063)	−0.475*** (0.045)	−0.079 (0.074)	−0.054 (0.081)	0.197** (0.066)

Notes: DISR = discruptive/antisocial, SELF = self-absorbed, COM = communiation disturbance, ANX = anxiety, SOC = social relating. GAC = general adaptive composite score (percentile rank). Coding: gender 0 = female, 1 = male. [a]Overall assessment based on Item 94 of the German DBC-T (full version). *$p < .05$. **$p < .01$. ***$p < .001$.

self-absorbed factor, no effect was observed (0.060, $p = .101$). Age showed a small negative effect on disruptive/antisocial behaviour, on self-absorbed behaviour, and on communication disturbances at T1. No effects were found for anxiety and social relating. At T2, the results followed the same pattern.

For the percentile rank of the GAC, negative effects were observed across all DBC-T subscales, indicating the lower a student's general adaptive skills, the greater their emotional and behavioural problems. Small negative effects showed for disruptive/antisocial behaviour, communication disturbances, anxiety, and social relating; in contrast, medium to large negative effects were reported for self-absorbed behaviour (T1: −0.619, $p < .001$; T2: −0.747, $p < .001$). The effect sizes for the three domains of adaptive behaviour varied considerably across the five subscales of the DBC-T. Conceptual skills only showed a small negative effect on communication disturbances, both at T1 (−0.233, $p < .001$) and T2 (−0.226, $p < .001$). Social skills had a medium-sized negative effect on self-absorbed behaviour and social relating at T1. A similar picture emerged at T2, however with an additional small negative effect on anxiety (−0.144, $p < .05$). Regarding practical skills, small negative effects were reported for disruptive/antisocial behaviour, social relating, and communication disturbances, but the latter only occurred at T1. The negative effect for self-absorbed behaviour was medium in effect size and somewhat more pronounced at T2 (−0.475, $p < .001$).

All five latent factors in the ESEM MIMIC model at T2 could be predicted by the overall rating of behavioural and emotional problems measured with item 94 of the DBC-T at T1. The significant positive effects were small in effect size, ranging from 0.22 for anxiety to 0.36 for disruptive/antisocial behaviour and social relating.

Discussion

In this study, we evaluated the factorial structure, internal consistency, and multiple aspects of validity of the German DBC-T. Our evaluation was based on a large sample of young people aged 4–19 years and their special needs teachers or other school staff members, who completed the DBC-T twice within a period of 7–9 months.

Regarding the factorial structure, our findings confirmed that the German DBC-T is best represented by five factors, consistent with the revised English and Dutch DBC versions (Dekker et al., 2002a, 2002b) and as expected for the German DBC-T (Einfeld et al., 2007). Our examination of the factor structure provides further support that the ESEM approach is suitable for analysing measurement instruments with complex data structure (see recommendation by Marsh et al., 2011), such as the DBC. In line with the construction of the DBC, our results indicated that the additional items (i.e., others) do not match well with the five original scales, and thus should only be applied for their intended use, namely the calculation of a total problem behaviour score (Einfeld et al., 2007).

The internal consistencies of the scales of the German DBC-T proved to be good at both measurement occasions. In contrast to earlier studies (Dekker et al., 2002b; Steinhausen & Winkler Metzke, 2004), even the internal consistency for the anxiety scale could be judged as good. One may argue that the reliability coefficients are not comparable, because we referred to McDonald's ω, whereas earlier studies employed Cronbach's α. Nonetheless, it should be emphasised that ω has been shown to provide a more accurate approximation of internal consistency, and that α corresponds in many cases to a lower bound of reliability (Dunn

et al., 2014; Sijtsma, 2009). Notwithstanding these differences, good reliability is particularly important for the DBC's application in practice.

With regard to the validity of the German DBC-T, the findings yielded evidence for the discriminant validity of the five subscales as indicated by moderate intercorrelations at both measurement occasions, suggesting that the five scales cover different but related aspects of emotional and behavioural problems. Furthermore, the discriminant validity of the DBC-T scales was underlined by small negative associations with the general score for adaptive behaviour, except for the self-absorbed scale, which showed moderate to large negative associations. Focusing on the three domains of adaptive behaviour, the level of social and practical skills, but not of conceptual skills, was related to self-absorbed behaviour. As expected, the social skills domain of adaptive behavior was also moderately associated with social relating as measured by the DBC-T, which, to some extent, could be interpreted as convergent validity. Nevertheless, the predominantly small and in several cases non-significant associations with the adaptive behaviour domains support discriminant validity for the DBC-T, consistent with the findings of Dekker et al. (2002b). Likewise, self-absorbed behaviour was moderately associated with general adaptive behaviour in both studies. It should be noted that Dekker et al. (2002b) employed the Dutch VABS to assess adaptive behaviour, while we relied on a German translation of the ABAS-3, and that the three domains of both instruments do not completely overlap.

The validity of the DBC-T was further substantiated by the results concerning students' gender and age, which pointed to appropriate known group validity. Gender generally had a small to medium effect on the scores of the five DBC-T scales. As to be expected, boys scored higher than girls, except for on the anxiety scale. This exception is in accordance with the general predominance of internalising problems in girls relative to boys (Rescorla et al., 2007). Interestingly, no gender-related effect was found for self-absorbed behaviour at T2. In line with previous studies (e.g., Chandler et al., 2015; Einfeld et al., 2006), problem behaviour decreased slightly with increasing age, that is across the range of 4–19 years. More specifically, a small negative effect of age was found for disruptive/antisocial behaviour, self-absorbed behaviour, and communication disturbances, whereas for anxiety and social relating, no effect was observed.

In a more exploratory manner, we additionally checked the predictive value of the overall rating of a child's behavioural and emotional problems assessed with the last item of the DBC-T. At T2, all five subscale scores could be predicted by the overall rating at T1. Considering the small effect sizes, the predictive value of item 94 should be estimated as minor.

To conclude, some limitations and future directions should be addressed. First and foremost, no direct measure of intellectual functioning was available in our study. While clinical diagnosis of ID was ensured by considering a sample attending specialised school setting, future studies might incorporate more detailed information on intellectual functioning, for instance, in order to standardise the German DBC-T based on ID severity. In view of the increasing importance of adaptive behaviour as assessment criteria for ID (e.g., Tassé, 2013), adaptive behaviour might be considered jointly with intellectual functioning. Considering that some disorders such as autism spectrum disorders can impact adaptive behaviour, independently or partly independently from ID severity, corresponding information could be used as covariate.

Second, the ESEM analyses pointed to several substantial cross-loadings. Furthermore, some items had very low observed frequencies. While the inclusion of all items is being advocated because of their clinical relevance, one might consider developing a shorter DBC-T version as has happened for the English DBC-P. Covering 24 items, the short form DBC-P24 (Taffe et al., 2007) has proved suitable for estimating a mean behaviour problem score, providing a brief and highly sensitive measure for research purposes.

Third, we only accounted for students' gender, age, and adaptive behaviour when investigating the construct validity of the German DBC-T. Future studies might benefit from including further characteristics (e.g., need for care, language competences) or from investigating additional aspects of validity, such as the German DBC-T's convergent validity with other questionnaires (e.g., CBCL) or between different informants (e.g., teacher and parent reports). Previous studies have indicated a bias in teachers' judgement accuracy of students' emotional and social characteristics, which was related to the presence or absence of a student's special educational needs status (Schwab et al., 2020; Venetz et al., 2019). Therefore, it would be worthwhile to investigate for possible bias in teachers' judgment accuracy in the application of the DBC-T and to consider the role of, for instance, teachers' professional experience, attitudes towards intellectual disability or their stress levels related to problem behaviours.

Fourth, it should be noted that we did not consider the alternative scales based on the DBC that are related to mental disorders (e.g., depression, hyperactivity). Further investigations could focus on a teacher version of the German autism screener (Steinhausen & Winkler

Metzke, 2004), which was developed following the DBC Autism Screening Algorithm (DBC-ASA; Brereton et al., 2002).

Fifth, our findings on age were cross-sectional, even if we examined age-related effects at both measurement occasions. Besides testing for one-year stability, future studies could investigate development over time. Finally, it would be worthwhile to account for the context specificity of a child's behaviour problems (Dworschak et al., 2016). Since behavioural and emotional problems are not merely stable dispositions but vary across different situations, it might be helpful for future studies to assess a person's behavioural and emotional reactions in situ, meaning in different momentary social contexts (Zurbriggen et al., 2018).

Taken together, our findings provide the support that the German DBC-T yields good psychometric properties, which constitute essential prerequisites for its use in research and practice. Furthermore, our study offers suggestions for additional evaluation and development of the DBC generally and the German DBC-T particularly. Having said that, further studies are warranted addressing more content-related issues, such as the context-dependency of behavioural and emotional problems in children and adolescents with intellectual disabilities, and the adequate use of standardised instruments in practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This article is based on data from a project funded by the Swiss National Science Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung) [grant number SNF-172773].

ORCID

Carmen L. A. Zurbriggen http://orcid.org/0000-0002-0906-2000
Christoph M. Müller http://orcid.org/0000-0001-6921-1322

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Department of Psychiatry, University of Vermont.

American Association on Intellectual and Developmental Disabilities (Ed.). (2021). *Intellectual disability: Defnition, classifcation, and systems of support* (12th ed.). AAIDD.

Amstad, M., & Müller, C. M. (2020). Students' problem behaviours as sources of teacher stress in special needs school for individuals with intellectual disabilities. *Frontiers in Education*, *4*, 159. https://doi.org/10.3389/feduc.2019.00159

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Bienstein, P., Döpfner, M., & Sinzig, J. (2017). Fragebogen zu den Alltagskompetenzen: ABAS-3. Englische Fassung: Patti L. Harrison & Thomas Oakland. Deutsche Evaluationsfassung [Questionnaire on adaptive behavior: ABAS-3. English Version: Patti L. Harrison & Thomas Oakland]. TU Dortmund.

Brereton, A. V., Tonge, B. J., MacKinnon, A. J., & Einfeld, S. L. (2002). Screening young people for autism with the Developmental Behavior Checklist. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(11), 1369–1375. https://doi.org/10.1097/00004583-200211000-00019

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Buckley, N., Glasson, E. J., Chen, W., Epstein, A., Leonard, H., Skoss, R., Jacoby, P., Blackmore, A. M., Srinivasjois, R., Bourke, J., Sanders, R. J., & Downs, J. (2020). Prevalence estimates health problems in children and adolescents with intellectual disabilities: A systematic review and meta-analysis. *Australian & New Zealand Journal of Psychiatry*, *54*(10), 970–984. https://doi.org/10.1177/0004867420924101

Chandler, S., Howlin, P., Simonoff, E., O'Sullivan, T., Tseng, E., Kennedy, J., Charman, T., & Baird, G. (2015). Emotional and behavioural problems in young children with autism spectrum disorder. *Developmental Medicine & Child Neurology*, *58*(2), 202–208. https://doi.org/10.1111/dmcn.12830

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

de Ruiter, K. P., Dekker, M. C., Douma, J. C. H., Verhulst, F. C., & Koot, H. M. (2008). Development of parent- and teacher-reported emotional and behavioural problems in young people with intellectual disabilities: Does level of ID matter? *Journal of Applied Research in Intellectual Disabilities*, *21*(8), 70–80. https://doi.org/10.1111/j.1468-3148.2007.00370.x

de Ruiter, K. P., Dekker, M. C., Verhulst, F. C., & Koot, H. M. (2007). Developmental course of psychopathology in youths with and without intellectual disabilities. *Journal of Child Psychology and Psychiatry*, *48*(5), 498–507. https://doi.org/10.1111/j.1469-7610.2006.01712.x

Dekker, M. C., Nunn, R. J., Einfeld, S. E., Tonge, B. J., & Koot, H. M. (2002a). Assessing emotional and behavioral problems in children with intellectual disability: Revisiting the factor structure of the Developmental Behavior Checklist. *Journal of Autism and Developmental Disorders*, *32*(6), 601–610. https://doi.org/10.1023/A:1021263216093

Dekker, M. C., Nunn, R. J., & Koot, H. M. (2002b). Psychometric properties of the revised Developmental Behaviour Checklist scales in Dutch children with intellectual disability. *Journal of Intellectual Disability Research*, *46*(1), 61–75. https://doi.org/10.1046/j.1365-2788.2002.00353.x

Dekker, M. C., Koot, H. M., van der Ende, J., & Verhuls, F. C. (2002c). Emotional and behavioral problems in children and adolescents with and without intellectual disabilities. *Journal of Child Psychology and Psychiatry*, *43*, 1087–1098. https://doi.org/10.1111/1469-7610.00235

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Dworschak, W., Ratz, C., & Wagner, M. (2016). Prevalence and putative risk markers of challenging behaviorin students with intellectual disabilities. *Research in Developmental Disabilities*, *58*, 94–103. https://doi.org/10.1016/j.ridd.2016.08.006

Einfeld, S. L., Piccinin, A. M., Mackinnon, A., Hofer, S. M., Taffe, J., Gray, K. M., Bontempo, D. E., Hoffman, L. R., Parmenter, T., & Tonge, B. J. (2006). Psychopathology in young people with intellectual disability. *JAMA*, *296*(16), 1981–1989. https://doi.org/10.1001/jama.296.16.1981

Einfeld, S. L., & Tonge, B. J. (1992). *Manual for the developmental behaviour checklist*. Clayton, Melbourne and Sydney: Monash University for Developmental Psychiatry and School of Psychiatry, University of New South Wales.

Einfeld, S. L., & Tonge, B. J. (1995). The Developmental Behaviour Checklist: The development and validation of an instrument to assess behavioural and emotional disturbance in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, *25*(2), 81–104. https://doi.org/10.1007/BF02178498

Einfeld, S. L., & Tonge, B. J. (2002). *Manual for the Developmental Behaviour Checklist, Primary Carer Version (DBC-P) & Teacher Version (DBC-T)* (2nd ed.). Monash University Centre for Developmental Psychiatry, Australia.

Einfeld, S. L., & Tonge, B. J. (n.d.). Developmental Behaviour Checklist. Information package. http://www.med.monash.edu.au/assets/docs/scs/psychiatry/dbc-info-package.pdf

Einfeld, S. L., Tonge, B. J., & Parmenter, T. (1999). *Manual for the Developmental Behaviour Checklist – Teacher Version*. Monash University Centre for Developmental Psychiatry, Australia.

Einfeld, S. L., Tonge, B. J., & Steinhausen, H.-C. (2007). *VFE: Verhaltensfragebogen bei Entwicklungsstörungen: Deutsche Version der Developmental Behavior Checklist (DBC)* [VFE: Behavioral questionnaire for developmental disorders: German version of the Developmental Behavior Checklist]. Hogrefe.

Emerson, E., & Einfeld, S. L. (2011). *Challenging behaviour*. Cambridge University Press.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Gray, K., Tonge, B. J., Einfeld, S. L., Gruber, C., & Klein, A. (2018). *DBC-2. Developmental Behavior Checklist 2*. Western Psychological Services.

Harrison, P. L., & Oakland, T. (2015). *ABAS-3: Adaptive behavior assessment system – third edition*. Western Psychological Services.

Hastings, R. P. (2002). Parental stress and behaviour problems of children with developmental disability. *Journal of Intellectual and Developmental Disability*, *27*(3), 149–160. https://doi.org/10.1080/1366825021000008657

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Li, C.-H. (2015). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu und Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–119. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, T. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220

Marsh, H. W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, *103*(3), 701–732. https://doi.org/10.1037/a0024122

Marsh, H. W., Vallerand, R. J., Lafrenière, M.-A. K., Parker, P., Morin, A. J. S., Carbonneau, N., Jowett, S., Bureau, J. S., Fernet, C., Guay, F., Salah Abduljabbar, A., & Paquet, Y. (2013). Passion: Does one scale fit all? Construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychological Assessment*, *25*(3), 796–809. https://doi.org/10.1037/a0032573

McDonald, R. P. (2013). Modern test theory. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods. Volume 1: Foundations* (pp. 118–143). Oxford University Press.

McKenzie, K., Milton, M., Smith, G., & Ouellette-Kuntz, H. (2016). Systematic review of the prevalence and incidence of intellectual disabilities: Current trends and issues. *Current Developmental Disorders Reports*, *3*(2), 104–115. https://doi.org/10.1007/s40474-016-0085-7

Molteno, G., Molteno, C. D., Finchilescu, G., & Dawes, A. R. L. (2001). Behavioural and emotional problems in children with intellectual disability attending special schools in Cape Town, South Africa. *Journal of Intellectual Disability Research*, *45*(6), 515–520. https://doi.org/10.1046/j.1365-2788.2001.00368.x

Morin, A. J. S., Maïano, C., Scalas, L. F., Aşçı, F. H., Boughattas, W., Abid, S., Mascret, N., Kara, F. M., Fadda, D., & Probst, M. (2018). Cross-cultural validation of the short form of the Physical Self Inventory (PSI-S). *Sport, Exercise, and Performance Psychology*, *7*(1), 60–79. https://doi.org/10.1037/spy0000096

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Müller, C. M., Cillessen, A. H. N., Egger, S., & Hofmann, V. (2021). Peer influence on problem behaviors among students with intellecutal disabilities. *Research in Developmental Disabilities*, *114*, 103994. https://doi.org/10.1016/j.ridd.2021.103994

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait-multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, *59*(1), 195–213. https://doi.org/10.1348/000711005X67490

Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., Bird, H., Chen, W., Dobrean, A., Döpfner, M., Erol, N., Fombonne, E., Fonseca, A., Frigerio, A., Grietens, H., Hannesdottir, H., Kanbayashi, Y., Lambert, M., Larsson, B., … Verhulst, F. C. (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders*, *15*(3), 130–142. https://doi.org/10.1177/10634266070150030101

Rescorla, L. A., Bochicchio, L., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., Bilenberg, N., Bird, H., Dobrean, A., Erol, N., Fombonne, E., Fonseca, A., Frigerio, A., Fung, D., Lambert, M., Leung, P., Liu, X., Markovic, I., … Verhulst, F. C. (2014). Parent-teacher agreement on children's problems in 21 societies. *Journal of Clinical Child and Adolescent Psychology*, *43*(4), 627–642. https://doi.org/10.1080/15374416.2014.900719

Rice, L. J., Emerson, E., Gray, K. M., Howlin, P., Tonge, B. J., Warner, G. L., & Einfeld, S. L. (2018). Concurrence of the strengths and difficulties questionnaire and developmental behaviour checklist among children with an intellectual disability. *Journal of Intellectual Disability Research*, *62*(2), 150–155. https://doi.org/10.1111/jir.12426

Schwab, S., Zurbriggen, C. L. A., & Venetz, M. (2020). Agreement among student, parent and teacher ratings of school inclusion: A multitrait-multimethod analysis. *Journal of School Psychology*, *82*, 1–16. https://doi.org/10.1016/j.jsp.2020.07.003

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales: Interview edition, survey form manual*. American Guidance Service.

Steinhausen, H.-C., & Winkler Metzke, C. (2004). Differentiating the behavioural profile in autism and mental retardation and testing of a screener. *European Child & Adolescent Psychiatry*, *13*(4), 214–220. https://doi.org/10.1007/s00787-004-0400-4

Steinhausen, H.-C., & Winkler Metzke, C. (2005). Der Verhaltensfragebogen bei Entwicklungsstörungen (VFE). Psychometrische Kennwerte und Normierung [The Developmental Behavior Checklist (DBC) – German version. Psychometric properties and norms]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *34*(4), 266–276. https://doi.org/10.1026/1616-3443.34.4.266

Steinhausen, H.-C., & Winkler Metzke, C. (2011). Der Verhaltensfragebogen bei Entwicklungsstörungen im Erwachsenenalter (VFE-ER). Psychometrische Kennwerte und Normierung [The Developmental Behaviour Checklist for Adults – German version. Psychometric properties and norms]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *40*(3), 160–171. https://doi.org/10.1026/1616-3443/a000094

Taffe, J. R., Gray, K. M., Einfeld, S. L., Koot, H. M., Emerson, E., Koskentausta, T., & Tonge, B. J. (2007). Short form of the Developmental Behaviour Checklist. *American Journal of Mental Retardation*, *112*(1), 31–39. https://doi.org/10.1352/0895-8017(2007)112[31:SFOTDB]2.0.CO;2

Tassé, M. J. (2013). Adaptive behavior. In M. L. Wehmeyer (Ed.), *The Oxford handbook of positive psychology and disability* (pp. 105–115). Oxford University Press.

Tóth-Király, I., Bõthe, B., Rigó, A., & Orosz, G. (2017). An illustration of the exploratory structural equation modeling (ESEM): framework on the passion scale. *Frontiers in Psychology*, *8*, 1968. https://doi.org/10.3389/fpsyg.2017.01968

Venetz, M., Zurbriggen, C. L. A., & Schwab, S. (2019). What do teachers think about their students' inclusion? Consistency of students' self-reports and teacher ratings. *Frontiers in Psychology*, *10*, 1637. https://doi.org/10.3389/fpsyg.2019.01637

Yuan, K. H., & Bentler, P. M. (2004). On chi-square difference and z-tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, *64*(5), 737–757. https://doi.org/10.1177/0013164404264853

Zurbriggen, C. L. A., Schwab, S., de Boer, A., & Koglin, U. (2018). Assessing behavior difficulties in students (editorial). *European Journal of Psychological Assessment*, *34*(2), 65–68. https://doi.org/10.1027/1015-5759/a000468