Development and Validation of a Multi-Informant Measure of Social Behaviors for Youth with Intellectual Disabilities

Elizabeth Olivier*, Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada
Alexandre J.S. Morin*, Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada
Danielle Tracey, School of Education, Western Sydney University, Sydney, Australia
Neha Verma, Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada
Céleste Dubé, Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada
Cynthia Gagnon, Département de psychoéducation et de psychologie, Université du Québec en Outaouais, Saint-Jérome, Canada
Rhonda G. Craven, Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia
Christophe Maïano, Cyberpsychology Laboratory and Département de psychoéducation et de psychologie, Université du Québec en Outaouais, Saint-Jérome, Canada

* The order of appearance of the first and second authors (E.O. and A.J.S.M.) was determined at random: Both should be considered first authors.

Corresponding author:
Alexandre J.S. Morin,
Department of Psychology, Concordia University, Canada,
7141 Sherbrooke W, Montreal, QC, Canada, H4B 1R6
alexandre.morin@concordia.ca

Abstract
Youth with intellectual disability (ID) are at an increased risk of displaying fewer prosocial behaviors and more numerous aggressive behaviors in various environments. This study proposes a new multi-informant (youth, teachers, and parents) measure of social behaviors for youth with ID. The sample includes 348 youth with mild (51.41%) and moderate (48.59%) levels of ID, aged 11-22 years old (*M*=15.73, *SD*=2.14; including 138 females), enrolled in secondary schools in Canada (French-speaking; *N*= 116; 33.33%) and Australia (English-speaking; *N*= 232; 66.67%). Measures were completed by the participants, their teachers, and their parents. Results support the reliability, factor validity, discriminant validity (in relation to sex, ID level, and country), concurrent validity (with measures of victimization, depression, hyperactivity-inattention), and one-year test-retest stability of the measure. Youth, teachers, and parents all provided a complementary perspective on youth social behaviors, consistent with youth adjusting their behaviors to the various environments in which they share social interactions.

Keywords: Measurement; social behaviors; prosocial skills; aggressive behavior; intellectual disability; special education needs.

        Youth who demonstrate prosocial considerations and refrain from being aggressive toward others develop more positive social relationships with members from the social systems with whom they regularly interact (adult caregivers, peers, siblings, etc.; Fox & Boulton, 2005; Hubbard et al., 2010). Compared to typically developing (TD) youth, youth with intellectual disabilities (ID) tend to display fewer prosocial behaviors and more numerous aggressive behaviors (Bailey et al., 2019). Prosocial behaviors contribute to the quality of life and positive post-school outcomes for youth with ID, whereas problematic behaviors such as aggression have a detrimental effect on their quality of life (Balboni et al., 2020; Dell'Armo & Tassé, 2019). The social nature of these behaviors makes them susceptible to change as a function of the environment (Aarts et al., 2003; Hubbard et al., 2010), which highlights the possibility of targeting these behaviors as part of preventive interventions. These contextual variations highlight the need to assess these behaviors in the context in which they arise naturally, such as home and school. This assessment is made all the more difficult by the rarity of measures available to support researchers interested in the social behaviors of youth with ID, in part due to the challenges associated with the development of proper self-report measures among this population. As a result, youth's voices and critical perspectives on their own behaviors are often neglected in research. These observations have led to calls for the development and validation of comprehensive measures of social behaviors for youth with ID (May & Kennedy, 2010). This study addresses this need by proposing a multi-informant (youth, parents, and teachers) measure of social behaviors designed for youth with ID.

Role of Social Behaviors for Youth with ID

        Youth who display positive social behaviors show prosocial behaviors and refrain from adopting aggressive behaviors in their social surroundings (Fox & Boulton, 2005; Hubbard et al., 2010). According to the APA (2020), prosocial behaviors are a subset of social skills denoting altruism and aiming to benefit others (e.g., helping, being considerate, sharing). Aggression, and more specifically verbal and physical aggression, involves behaviors aiming to psychologically or physically hurt others (APA, 2020), either in reaction to a frustration (reactive aggression) or by provocation (proactive aggression) (Salmivalli & Nieminen, 2002). Aggression includes behaviors such as saying bad things, scaring, pinching, hitting, or attacking others.

        Among TD youth, research on the joint development of prosocial and aggressive behaviors is fairly common. In this population, although prosocial youth are less likely to be aggressive and vice versa, prosocial and aggressive behaviors are not mutually exclusive. For instance, some may display a strategic combination of aggression and prosociality with their peers to achieve a better social status (Hartl et al., 2020). Despite this interconnected nature, few studies conducted among youth with ID have considered the joint role of prosocial and aggressive behaviors. Compared to TD youth, youth with ID tend to display fewer prosocial behaviors and more aggressive ones (Bailey et al., 2019). Impairments in their cognitive flexibility and ability to shift their attention quickly could explain why youth with ID might fail to adjust their behavior adequately across social situations and may misread social information. These youth may react more aggressively and in a less prosocial manner to situations where such reactions are not appropriate (Visser et al., 2015). However, these studies also suggest that the social behaviors of youth with ID can improve, perhaps even more than that of their TD peers, as a result of environmental influences (e.g., Bailey et al., 2019; Kurtek, 2018). In school, aggression may be used to thrive in a competitive environment, whereas prosociality may be more adapted to the home environment. Kurtek (2018) also suggests that aggression might be used to gain autonomy from parents.

        These observations all highlight the need to improve our understanding of the joint role of prosocial and aggressive behaviors in different environments among youth with ID. The school and home environments are those among which youth with ID have the closest and more numerous social interactions. The scarcity of studies assessing these behaviors among youth with ID seems to stem from the lack of validated measures: (1) jointly assessing these two types of social behaviors; and (2) doing so in a way that captures the unique perspective of the youth who is the target of these measures. Most studies rely on informant reports, usually a parent, a teacher, or a clinician. Although informative, these reports do not represent the full range of youth social behaviors given that many prosocial or aggressive interactions among peers tend to appear when adults cannot observe them (e.g., school recess; Bradshaw et al., 2007; Craig et al., 2000). As the gold standard in assessing youth behavior is to combine various sources of information including self-reports and informant reports (Huang, 2017; De Los Reyes, 2011; Renk, 2005), this study aims to assess the perspectives of youth, their teachers, and their parents on

social behaviors.

Measuring Social Behaviors among Youth with ID

Over the years, different scales have been developed to measure prosocial or aggressive behaviors separately among samples of participants with ID. From a psychometric perspective, each of these scale present their own strengths and weaknesses. First, the Dynamic Risk Outcome Scale (e.g., Delforterie et al., 2018), the Staff Observation Aggression Scale – Revised (e.g., van den Bogaard et al., 2018), and the Modified Overt Aggression Scale (e.g., Oliver et al., 2007) all include components assessing both prosocial skills and aggressive behaviors. However, these scales are mainly designed to help therapists and practitioners measure the progress of their patients, which is not necessarily suitable to a more global evaluation of social interactions occurring in their natural non-therapeutic environments, such as the school or the home. Second, the Behavior Problem Inventory for individuals with ID (e.g., Rojahn et al., 2012a, 2012b) and the Behavioral Assessment Scale for Indian Children with Mental Retardation (e.g., used by Lahkan et al., 2018) both rely on informant (e.g., parent or caregiver) reports of verbal and physical aggression. However, both scales are very long (respectively 52 – 30 in the short version – and 71 items), which makes them unsuitable for large scale studies assessing numerous aspects of youth's functioning. Third, the Social Performance Survey Schedule (encompassing prosocial behaviors; e.g., Rojahn et al., 2011) and the Psychopathology Checklist for Adults with ID (encompassing aggressive behaviors; e.g., Hove & Havik, 2008) have only been validated for adult populations, and only the former includes a self-report version, whereas the latter requires college-educated informants. Fourth, the Diagnostic Adaptive Behavior Scale (e.g., Tassé et al., 2019) and the Vineland Adaptive Behavior Scale (e.g., Sparrow et al., 2016) both include a social skills component. These measures are rated by an informant who knows the youth with ID well (typically a parent or teacher), but do not include a self-reported questionnaire. Finally, the Aberrant Behavior Checklist (e.g., Rojahn et al., 2011) includes aggressive behaviors and is available for different informants (including parents and teachers), but again does not include a self-reported questionnaire.

In summary, none of these scales simultaneously: (a) encompass prosocial and aggressive behaviors; (b) include a validated self-report component suitable for youth with ID; or (c) include matching sets of items that can be used across youth's, parents', and teachers' reports. For these reasons, it is not uncommon for researchers who wish to capture youth perceptions (e.g., Visser et al., 2015) or both components of social behaviors (e.g., Bailey et al., 2019) to rely on instruments that have never been validated among a population with ID. The present study seeks to fill this gap by proposing, and validating, a multi-informant measure of social behaviors suitable for youth with ID, and short enough to be applicable to both research and clinical settings. More precisely, we propose a multi-informant (parents, teachers, and youth) adaptation of the prosocial behavior scale from the Strengths and Difficulties Questionnaire (Goodman, 2001; Capron et al., 2007) and of the deviant behavior scale from the Children's Behavior Questionnaire (Quebec Longitudinal Study of Child Development, 2006, 2008; Tremblay et al., 1987) suitable for the assessment of social behaviors among youth with ID. The two instruments from which these scales have been adapted have the advantages of: (a) being largely used among TD youth; (b) being relatively short; (c) having already been validated in English and French, and (d) already including youth, parent, and teacher versions.

The Present Study

Scale Development and Validation

The present study seeks to develop and validate a multi-informant (youth, teachers, and parents) measure of prosocial and aggressive behaviors specifically designed for youth with ID. Furthermore, by capitalizing on a cross-cultural sample of English-speaking Australian youth and French-speaking Canadian youth, we were able to simultaneously develop, and validate, the resulting measure in the English and French languages. In doing so, we relied on items taken from two measures well-established for research on TD populations, the Strength and Difficulties Questionnaire (Goodman, 2001; Capron et al., 2007) and the Children's Behavior Questionnaire (Tremblay et al., 1987) for which self-, parents-, and teacher-reports already exist in English and French. We expected youth, teachers, and parents to reliably assess the prosocial and aggressive behaviors displayed by youth with ID. We also expected all three raters to provide complementary perspectives (i.e., distinct, with r<.500) consistent with the idea that youth may adjust their behaviors to different settings, hold unique views compared to other informants, and reinforcing the need to incorporate multiple-informants to obtain a more nuanced understandings of the reality.

Generalizability across Youth Characteristics

Our second objective is to ascertain that the psychometric properties of this new instrument would remain unchanged (i.e., would function equally well) as a function of various youth characteristics (sex, ID level, age, and country/language). This verification involves tests of measurement invariance (Millsap, 2011), conducted to verify whether participants' characteristics affect (i.e., bias) the respondents' pattern of responses to specific items over and above the effects of these characteristics on the latent constructs. Consistent with our expectation that the resulting instrument will be generalizable to all types of youth with ID, we expect items to function in the same manner regardless of sex, ID level, age, and country/language. In addition, these tests of measurement invariance also make it possible to assess the discriminant validity of the resulting measure by verifying whether and how these characteristics (i.e., sex, ID level, and country/language) are associated with mean differences at the levels of the latent constructs (i.e., prosocial and aggressive behaviors). In these tests, discriminant validity can be demonstrated when observed mean differences matched those typically observed in previous research. In this regard, studies conducted among populations with ID generally find that males and females tend to display the same level of prosocial and aggressive behaviors (Hove et al., 2010; Lakhan & Kishore, 2018), leading us to expect a lack of sex-related differences. In contrast, some studies suggest that youth with more severe levels of ID are more likely to be aggressive (e.g., Lakhan & Kishore, 2018) and to display poorer social relationships (e.g., Totsika et al., 2014), thus leading us to expect higher levels of aggressive behaviors and lower levels of prosocial behaviors among youth with moderate, relative to mild, levels of ID. Finally, there should be no effect of country/language on the constructs considered here, consistent with the full equivalence of the linguistic versions of the instrument and the similarity in the general life and educational conditions of people with ID observed in Canada and Australia.

Concurrent Validity

Our third objective was to assess the concurrent validity of this new instrument by testing associations between youth, teacher, and parental ratings of youth's social behaviors with their ratings of depressive symptoms, victimization, and hyperactivity-inattention. The assessment of the concurrent validity (also referred to as convergent or criterion-related validity) of a new measure has long been acknowledged as a core component of psychometric validation studies (e.g., Boateng et al., 2018). Tests of concurrent validity verify whether a newly developed measure shares associations with other variables that correspond to the associations generally reported in research between the constructs assessed by this new measure (i.e., social behaviors) and these other variables (Boateng et al., 2018; Furr & Bacharach, 2014; Price 2017; Raykov & Marcoulides, 2011). Based on existing studies conducted among youth or adults with ID, participants presenting higher levels of aggressive behaviors should present a higher risk of victimization (Clark et al., 2016), depression (Davies & Oliver, 2014), and hyperactivity-inattention (Reiter & Lapidot-Lefler, 2007). Although fewer studies have focused on prosocial behaviors, some indicate that, among individuals with ID, prosocial skills predict a lower risk of depression (Hartley & Birgenheir, 2009) and victimization (Reiter & Lapidot-Lefler, 2007). Moreover, consistent with the idea that each informant – youth, teacher, or parent – has a unique and complementary perspective, we expect stronger associations between ratings of social behaviors and the three outcomes when evaluated by the same rater (Turk et al., 2012).

Test-retest Stability

Our fourth objective was to investigate if the psychometric properties of the instrument would be replicated one year later, as well as the test-retest stability of the ratings of each informant over the same period of time (e.g., Boateng et al., 2018; Furr & Bacharach, 2014; Price 2017; Raykov & Marcoulides, 2011). Because teachers change yearly, we expected their ratings to show a lower level of stability than those of youth and parents.

Method

Participants

This study is based on a sample of 348 students with mild (51.41%, corresponding to IQ scores between 50 and 69) and moderate (48.59%, corresponding to IQ scores 36 and 49) levels of intellectual disability (ID). ID classifications were determined using IQ scores, in line with the DSM-IV (APA, 2000) which was the official classification system used in the school records for participating students at the time of data collection. Youth were aged 11 to 22 years old (M=15.73, SD=2.14; including 138 females), enrolled in secondary schools in Canada (French-speaking; N= 116; 33.33%) and Australia

(English-speaking; *N*= 232; 66.67%). Furthermore, 88 (25.29%) of these youth had a reported comorbidity (i.e., 44 presented a comorbid autism spectrum disorder, 40 presented a comorbid genetic syndrome, and 4 both comorbid conditions). Teachers (n= 282; 119 in Canada and 163 in Australia) and parents (n= 179; 95 in Canada and 84 in Australia) also participated. Youth, teachers, and parents also participated at a second time point (T2). Of the youth, 240 (70 in Canada and 170 in Australia) were then retested one year later (37.92% females; 46.93% mild ID; 53.07% Moderate ID), as well as 146 of the teachers (45 in Canada and 101 in Australia) and 112 of the parents (63 in Canada and 49 in Australia).

Procedure

Participants were recruited within schools or community organizations that agreed to support this research. No compensation was offered for participation in Australia. Canadian participants were eligible to win one of 40 gift certificates ($30 CAD) annually. Parents (or legal representatives) of all participants actively provided signed informed consent for their own, and their child's participation. For parents of children recruited via participating schools, this consent form was directly sent to the parents by the school, together with an information letter, and the signed consent form was returned directly to the school where members of the research team collected it. Parents recruited outside of the participating schools received this material directly from the research team and returned the signed consent form to the researchers using a reply-paid envelope. All youth authorized to participate by their parents also had to actively provide their own written consent to participate.

The consent procedure granted the researchers access to school records, including youth's most recent level of intellectual functioning (only students with an official school-based ID classification were recruited). The Wechsler (2008) Intelligence Scale for Children – Fourth Edition (WISC-IV) was the IQ test most frequently used by the schools in both countries. When the last IQ assessment in the school records was older than four years, a new IQ assessment was conducted by a registered psychologist using the WISC-IV, the Wechsler Adult Intelligence Scale-IV (WAIS-IV), or the Leiter international performance scale-revised (Roid & Miller, 1997), depending on age and verbal ability. In Australia, a total of 34 participants were thus assessed by our research team, and all of those completed the Wechsler version corresponding to their chronological age (31 WISC-IV and 3 WAIS-IV). In Canada, 59 participants were re-assessed, 50 of those using the Wechsler version (21 WISC-IV or 29 WAIS-IV) corresponding to their chronological age, and 9 of them using the Leiter scale. This breakdown is not available for the majority of participants from whom we obtained IQ scores from the school records. We note that parents were informed that participating youth had to be able to understand the type of questionnaire items used in this study (irrespective of their levels of expressed verbal skills). As a result, non-verbal participants, and participants with comorbid conditions likely to interfere with their ability to complete the questionnaire, were self-excluded by their parents.

Participating students were met at their school (or at a time and location most convenient for the parents for participants recruited outside of schools) by members of the research team or trained research assistants who explained the goals and procedures of the study, as well as youth's right not to participate or to withdraw from the study without any consequences. Students were asked to actively and voluntarily consent to the study. Using sample questions, research assistants explained how to use the response scales (all involving graphical displays and pictograms). Testing was realized in small groups including up to 8 students with mild levels of ID or including 1 or 2 students with moderate levels of ID. A read-aloud assisted procedure was utilized to maximize understanding, and students were encouraged to ask questions if anything was unclear or if they failed to understand any questions. In addition, the research assistants routinely (more frequently with youth with moderate levels of ID) asked questions to verify youth's understanding of the questions. To help them answer youth's questions, the research assistants had access to a list of synonyms that they could use when youth did not understand a word. Sometimes, despite the available support, students remained unable to understand a question. Students were then instructed to select the "do not understand" option. In this study, this option was rarely selected. More precisely, on the 10 items rated by students to measure prosocial and aggressive behaviors, the lowest rate of DNU observed on an item was 0.29%, whereas the maximum rate of DNU observed for an item was 2.30%. This option was treated as a missing response. Details about the frequency at which items were not understood or left missing by the children are provided in Table S1 of the online supplements.

Parents of children enrolled in the targeted schools were asked to complete a questionnaire sent

to them by the school (or directly by the research team for those recruited outside of schools), each year of the study. Parents could complete the questionnaire at a time convenient for them, and return either to the schools or the researchers using a reply-paid envelope. Participating schools also agreed to distribute and collect teacher consent forms and questionnaires each year of the study. Teachers were encouraged to complete the questionnaire during the data collection process (questionnaires were recuperated by the research team), or at a time more convenient for them (questionnaires were sent to the research team using a reply-paid envelope). The study was approved by the research ethics committees of the Western Sydney University, Australian Catholic University, and Université du Québec en Outaouais.

Measures

Youth questionnaires were adapted for self-report completion by youth with ID following procedures similar to those used for the adaptation of other measures for this population (Maïano et al., 2009, 2011a, 2011b). Questions were maximally simplified, and items as well as response scales were associated with graphical depictions to facilitate understanding. This adaptation was realized through a collaborative process, including bilingual researchers familiar with this process and population, as well as teachers, psychologists, and psycho-educators all experienced in working with youth with ID. A first version of the adapted measures was pre-tested as part of a first pilot study conducted among youth (13 to 21 years old; n=8 in Canada and n=10 in Australia) with mild to severe ID, their teachers, and parents. This first pilot study was used to contrast different formulations of the questions and response scales, and alternative response format (verbal only, pictorial only, and combination). This initial pilot led to an improved version of the questionnaires (using a combination of graphical and verbal response scales). This improved version was trialed in a second pilot (n=6 youth in Canada and n=10 in Australia, teachers, and parents) to confirm the adequacy of the questionnaires and revise the final versions used in the main study. A more detailed presentation of the adaptation process can be consulted in the Scale Development section of the online supplements.

Social Behaviors. At both time points, youth, teachers, and parents rated a series of items drawn from the Strengths and Difficulties Questionnaire (Goodman, 2001; Capron et al., 2007) and the deviant behavior scales of the Children's Behavior Questionnaire (Quebec Longitudinal Study of Child Development, 2006, 2008; Tremblay et al., 1987) test battery, both available in French and English, to obtain a comprehensive coverage of prosocial and aggressive behaviors, conceptualized as two independent but correlated dimensions. These items were then adapted for youth with ID as part of the process described in the initial section of the online supplements. The verbal formulation and response scale used in the present study across informants is reported in Appendix A at the end of the online supplements, whereas the complete questionnaires are available free of charge from the corresponding author. Using the adapted measures, youth rated their prosocial (5 items; α_{t1}=.79; α_{t2}=.77; e.g., "You are attentive to other people's feelings") and aggressive (5 items; α_{t1}=.85; α_{t2}=.86; e.g., "You became physically aggressive when someone contradicted you") behaviors using a frequency scale ranging from 0 (*never*) to 5 (*5 times or more*). Parents and teachers answered a slightly longer set of matching scales: prosocial (7 items; e.g., "This student/My child is considerate of other people's feelings"; teachers: α_{t1}=.87; α_{t2}=.88; parents α_{t1}=.88; α_{t2}=.84) and aggressive (8 items; e.g., "This student/My child became physically aggressive when someone contradicted him/her"; teachers: α_{t1}=.91; α_{t2}=.94; parents α_{t1}=.89; α_{t2}=.88) behaviors exhibited by the target youth using a response scale ranging from 1 (*never*) to 5 (*very often*).

Peer Victimization. At both time points, youth, teachers, and parents were asked to report the frequency to which the target youth had been exposed to peer victimization using the relevant scale from the Socio-Educative Questionnaire (Janosz et al., 2007). These 14 items (youth: α_{t1}=.94; α_{t2}=.90) referred to acts of verbal (e.g., "Another student was rude or laughed at me"), physical (e.g., "Another student pushed, hit or kicked me"), and relational victimization (e.g., "Another student didn't want me to play with their friends") perpetrated by peers, and were rated on a frequency scale ranging from 0 (*never*) to 5 (*5 times or more*). Teachers and parents rated 14 matching items (teachers α_{t1}=.89; α_{t2}=.92; parents: α_{t1}=.91; α_{t2}=.90; e.g., "This student / My child has been called names or mean things by other students") using a frequency scale ranging from 1 (*never*) to 5 (*very often*).

Depressive Symptoms. At both time points, youth, teachers, and parents evaluated youth's levels of depressive symptoms using the Glasgow Depression Scale for people with intellectual disabilities (Cuthill et al., 2003). Youth completed the original version of the scale (21 items; α_{t1}=.89;

α_{t2}=.86; e.g., "Have you felt sad.") using a response scale ranging from 0 (*never*) to 4 (*always*). Teachers and parents completed the caregiver version of the same measure (16 items; teachers α_{t1}=.85; α_{t2}=.86; parents: α_{t1}=.81; α_{t2}=.77; e.g., "Has this student / My child appeared depressed?") using a response scale ranging from 1 (*never*) to 5 (*very often*).

Hyperactivity-inattention. At both time points, teachers and parents evaluated youth's levels of hyperactivity-inattention using five items (e.g., "This student / My child has difficulty staying on a task or completing work/activities.") from the Anxiety, Depression, and Mood Screen (ADAMS; Esbensen et al., 2003; Méthot & Morin, 2004), an instrument specifically developed for people with ID. These items were rated by teachers (α_{t1}=.80; α_{t2}=.85) and parents (α_{t1}=.74; α_{t2}=.70) using a severity scale ranging from 1 (*not a problem*) to 5 (*major problem*).

Covariates. Youth's sex (0=female; 1=male), country of residence (0=Canada; 1=Australia), age (11 to 22 years old), and ID level (0=mild; 1=moderate) were obtained via official school records.

Analyses

Model Estimation. All analyses were performed using Mplus 8.2 (Muthén & Muthén, 2019). Measurement models and tests of measurement invariance were conducted using the robust Weight Least Square with Mean and Variance adjusted statistics (WLSMV) estimator, which has been shown to outperform Maximum Likelihood (ML or ML robust) estimation procedures when using ordinal rating scales with asymmetric response thresholds such as those used in the present study (Finney & Di Stephano, 2013). All models were estimated using all of the available information, using missing data algorithms implemented in Mplus for WLSMV estimation (Asparouhov & Muthen, 2010). Measurement models based on youth self-reports obtained at T1 included 1.43% to 3.45% of missing data on the social behavior indicators (including the do not understand responses, see Table S1 of the online supplements) and 0.86% to 5.17% of missing responses on the outcome measures, whereas their T2 reports included 0.42% to 5.41% of missing data (social behavior indicators only). Measurement models involving teacher and parent ratings respectively included 0.00% to 2.48% and 0.56% to 1.68% of missing data at T1 on the social behavior indicators, 0.00% to 3.19% and 0.00% to 3.91% of missing data at T1 on the outcome measures, and 0.00% to 4.11% and 1.78% to 5.36% of missing data at T2 (social behavior indicators only).

Measurement Models Specification. We first contrasted alternative measurement models to establish the optimal representation of behavioral functioning separately for each informant (youth, teacher, and parent) at both time points. For comparison purposes, we first estimated a one factor solution. We then estimated the a priori two-factor solution, assuming that the prosocial items would be associated with a prosocial factor and that the aggressive items would be associated with an aggressive factor. Following recommendations formulated by Morin and colleagues (2016, 2020) for the assessment of multidimensional constructs, this a priori solution was estimated using four alternative measurement specification. We first relied on a confirmatory factor analytic (CFA) solution. In this solution, items were only allowed to define their a priori factor, factors were specified as correlated, and no cross-loadings was permitted. Second, we relied on an exploratory structural equation modeling (ESEM) representation. In this solution, factors were defined as in the CFA solution, but all cross-loadings were freely estimated between items and non-target factors but "targeted" to be as close to zero as possible using a confirmatory form of rotation procedure (i.e., target rotation; Morin et al., 2016, 2020). Third, we relied on a bifactor-CFA solution. In this solution, one global factor (G-factor) reflecting the variance shared among all items from the instrument, together with the assessment of two orthogonal specific factors (S-factors), defined as in the CFA solution, and reflecting subscale specificity left unexplained by the G-factor. Finally, we estimated a bifactor-ESEM solution, in which two orthogonal S-factors were defined as in the ESEM solution, and accompanied by the estimation of a G-factor.

This comparison was necessary based on statistical evidence highlighting the benefits of ESEM, revealing that freely estimating cross-loadings tends to result in more accurate estimates of factor correlations (Asparouhov et al., 2015) or regressions (Mai et al., 2018) as soon as cross-loadings as small as .10 are present. Furthermore, the estimation of bifactor solutions allowed us to explicitly test whether all behaviors included assessed in our measured could be conceptualized to form a single overarching continuum of prosocial behaviors (i.e., a G-factors on which prosocial and aggressive items are associated with factor loadings characterized by a different sign: positive/negative or negative/positive), while allowing us to obtain a direct estimate of the specificity remaining associated

with each separate dimension (Morin et al., 2016, 2020). Importantly, the comparison between a bifactor solution and the initial one-factor solution provided a direct test of whether this specificity is meaningful.

These alternative solutions were compared following Morin et al. (2016, 2020) recommendations stating that the first-order CFA and ESEM solutions need to be contrasted first, and that this comparison should favor the ESEM solution when it results in: "(a) an improved level of fit to the data; (b) reduced factor correlations; (c) small to moderate cross-loadings […]; (d) well-defined factors" (p.14). The retained solution should then be contrasted with its bifactor counterpart, which should be retained when it results in: "(a) an improved level of fit the data; (b) a well-defined G-factor; (c) at least some reasonably well-defined S-factors" (Morin et al., 2020, p. 14).

Measurement Invariance (differential item functioning, discriminant validity, and test-retest stability). After selecting the optimal solutions, we performed tests of measurement invariance (Millsap, 2011) as a function of sex, age (using a median split at 15.38 y.o.), country (linguistic version: Canada-French or Australia-English), level of ID (mild or moderate), and measurement point (T1 and T2), in the following sequence (Morin et al., 2011, Millsap, 2011): (i) configural invariance (same model with no other constraint); (ii) equal factor loadings (weak invariance); (iii) equal factor loadings and response thresholds (strong invariance); (iv) equal factor loadings, response thresholds, and item uniquenesses (strict invariance); (v) equal factor loadings, response thresholds, item uniquenesses, and latent variance-covariance matrix; (vi) equal factor loadings, response thresholds, item uniquenesses, latent variance-covariance matrix, and latent means. The last two steps are not required to establish invariance, but were tested for descriptive purposes. The most invariant model obtained across time points for each informant was then used to obtain estimates of test-retest correlations for each latent factor.

Concurrent Validity. The concurrent validity of the youth-, teacher-, and parent-rated social behaviors factors, was assessed in a series of structural equation models in which these latent factors were allowed to predict manifest scores on the outcome variables rated by youth (victimization and depression), as well as their teachers and parents (victimization, depression, and hyperactivity-inattention). Three models had to be estimated for youth-, teacher, and parent-ratings of social behaviors, but each model included all outcomes rated by all informants. These models accounted for age, sex, and level of ID. Variable correlations, together with descriptive statistics, are reported in Table S2 of the online supplements.

Model Fit. Model fit was assessed using the chi-square statistic (χ^2), the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI) (Hu & Bentler, 1999; Yu 2002). CFI and TLI values above .90 and .95 respectively support acceptable and excellent model fit. In contrast, RMSEA values smaller than .06, .08, and .10 respectively support excellent, good, and acceptable model fit. We also report the composite reliability of each factor using the omega coefficient (ω; McDonald, 1970), which has been recommended for CFA, ESEM, bifactor-CFA, and bifactor-ESEM (Morin et al., 2020). In tests of measurement invariance, increases in RMSEA of more than .02 and decreases in CFI and TLI of more than .01 reflect non-invariance (Chen, 2007). We also report WLSMV chi-square and chi-square difference tests, which were calculated using the Mplus DIFFTEST option (Muthén & Muthén, 2019).

Results

Measurement Models

Youth Self-Reports of Social Behaviors. The fit of the alternative models is reported in Table 1. Apart from the one-factor CFA solution which failed to attain an acceptable level of fit, the other solutions resulted in an excellent level of fit. Furthermore, the ESEM solution did not result in a meaningful improvement in model fit relative to the CFA solution. Likewise, the fit of the bifactor-ESEM solution was comparable to that of the bifactor-CFA solution, although both bifactor solutions resulted in an improved level of fit when compared to their first order counterparts. Thus, model fit information seems to support the superiority of a bifactor-CFA solution. However, as noted above, model fit is only one source of information to consider when comparing these models and should remain secondary to an examination of parameter estimates. When considering these various solutions, all solutions resulted in reasonably well-defined factors, both of the ESEM (first-order or bifactor) solutions revealed very few noteworthy cross-loadings, and the ESEM estimate of the factor correlation (r=-.37) was essentially identical to the CFA estimate (r=-.41). These observations appear to support

the superiority of a CFA, relative to ESEM, representation of the data, thus suggesting that it was not necessary to incorporate cross-loadings. Furthermore, the bifactor-CFA solution resulted in generally well-defined S- and G- factors, lending further support to its superiority relative to the first-order CFA solution.

Parameter estimates from this bifactor-CFA solution are reported in the top of Table 2. With two exceptions, these results revealed a well-defined G-factor ($|\lambda|=.33–.79$; $\omega=.87$) reflecting youth's global levels of social behavior problems (with negative factor loadings from the prosocial items and positive factor loadings from the aggressive items). They also revealed well-defined prosocial behavior ($|\lambda|=.32–.74$; $\omega=.73$), and aggressive behavior ($|\lambda|=.52–.83$; $\omega=.90$) S-factors. The exceptions were related to one item with a small and non-significant factor loading on the social behavior problems G-factor (i.e., *Play with friends*) and one item with a small and non-significant factor loading on its a priori S-factor (i.e., *Considerate*). The first of those items presented a much stronger loading on its a priori prosocial behavior S-Factor, whereas the second presented a much stronger loading on the social behavior problems G-factor. This solution was thus retained for further analyses.

Results from the tests of measurement invariance conducted on this solution are reported in Table S3 of the online supplements. These results first support the complete measurement invariance of this solution over time, across samples of boys and girls, and as a function of age. However, some indications of non-invariance were also evidenced as a function of youth's levels of ID and across countries. When we first consider youth's levels of ID, the analyses first failed to support the weak invariance of the solution. However, an examination of the parameter estimates obtained in the configural invariance model and of the modification indices associated with the model of weak invariance reveals that this non-invariance appears to be limited to three items. More precisely, the item "You forced others to give you something you wanted by scaring them" had a weaker factor loading ($\lambda=.30$) on the social behavior problems G-factor among youth with mild levels of ID than among youth with moderate levels of ID ($\lambda=.48$). Likewise, the item "You have been attentive to other people's feelings" had a weaker factor loading ($\lambda=.35$) on the prosocial behavior S-factor among youth with mild levels of ID than among youth with moderate levels of ID ($\lambda=.44$). Conversely, the item "You became physically aggressive or angry when someone hurt you" had a stronger factor loading ($\lambda=.85$) on the aggressive behavior S-factor among youth with mild levels of ID than among youth with moderate levels of ID ($\lambda=.75$). Once equality constraints on these three factor loadings were relaxed, the resulting model of partial weak invariance was supported by the data. The results also supported a model of partial strong invariance in which equality constraints has to be relaxed on three response thresholds as a function of youth levels of ID, suggesting that youth with moderate levels of ID had a slightly higher response threshold associated with moving from the first to the second response categories (from *never* to *1 time*) on the "You became physically aggressive or angry when someone hurt you" and "You became physically aggressive when something was taken away from you" items, but a slightly lower response threshold in moving from the third to the fourth (from *3 times* to *4 times*) response categories on the "You have been helpful when someone was hurt or feeling unwell" item, relative to their peers with mild levels of ID. The resulting model of partial strong invariance was supported, as well as the next models of strict, latent variance-covariance, and latent mean invariance.

When we turn our attention to tests of measurement invariance as a function of the country, which also corresponds to the two linguistic versions, the results first supported the weak invariance of the model, but not its strong invariance. In contrast, the results supported a model of partial strong invariance in which equality constraints has to be relaxed on three response thresholds associated with a single item ("You have been helpful when someone was hurt or feeling unwell"), suggesting that response thresholds associated with passing from the first (*never*) to the fourth (*3 times*) response categories were higher for Canadian-French, relative to Australian-English youth. From this model, the results also supported the strict invariance, and latent variance-covariance invariance of this solution across countries, but not its latent mean invariance. However, a model of partial latent mean invariance uncovered latent mean differences between countries suggesting that youth from Australia tended to score 1.17 *SD* higher than Canadian youth on the aggressive behavior S-factor. Scale score reliability estimates (i.e., Cronbach alpha) obtained within the various subgroups of youth with considered in this study are Table S4 of the online supplements.

Teacher Reports of Youth's Social Behaviors. The fit of the alternative models is reported in Table 1. Apart from the one-factor CFA solution which failed to attain a minimally acceptable level of

fit to the data, all of the alternative solutions resulted in an excellent level of fit to the data according to the CFI and TLI. In contrast, the RMSEA revealed a lack of fit for the first-order CFA and ESEM solutions, but an acceptable level of fit for their bifactor counterparts. These results thus support the value of adopting a bifactor solution. It is also interesting to note that, although the fit of the ESEM solution was higher than that of the CFA solution according to the CFI (ΔCFI=+.02), it was substantially lower according to the TLI (ΔTLI=-.02) and RMSEA (ΔRMSEA= +.02), suggesting a lack of parsimony. Parameter estimates from these solutions supported this interpretation in revealing very few noteworthy cross-loadings, generally well-defined factors across solutions, and ESEM estimates of the factor correlation (r=-.46) that is comparable to the CFA estimate (r=-.51). These considerations suggest that cross-loadings were not necessary, and even harmful. In contrast, the bifactor-CFA solution resulted in an improved, and acceptable, level of fit when compared to its CFA counterpart, and in well-defined factors. The parameter estimates from this bifactor-CFA solution are reported in the middle section of Table 2 and reveal a well-defined social behavior problems G-factor ($|\lambda|$=.25–.96; ω=.95) and aggressive behavior S-factor ($|\lambda|$=.63–.78; ω=.94). In contrast, the prosocial behavior S-factor was more weakly defined ($|\lambda|$=–.18–.80), although anchored in generally satisfactory factor loadings from a majority of items, as illustrated by a ω value of .76. This solution was retained for further analyses, and proved to be fully invariant as a function of youth's sex, age, ID level, and country (see Table S5 of the online supplements).

Parental Reports of Youth's Social Behaviors. The fit of the alternative models is reported in Table 1. Apart from the one-factor CFA solution which failed to attain a minimally acceptable level of fit to the data, all of the alternative solutions resulted in an excellent level of fit to the data according to the CFI and TLI. In contrast, the RMSEA revealed an acceptable level of fit to the data for the first-order CFA and ESEM solutions, but an excellent level of fit to the data for the bifactor solutions. These results thus again support the value of a bifactor solution. In addition, the ESEM solution only resulted in a minimal improvement in model fit relative to that of the CFA solution, and in a similar estimate of the factor correlation (ESEM r=-.21; CFA r=-.23). These considerations suggest, once again, that cross-loadings were not necessary to these solutions. When we turn our attention to the bifactor-CFA solution, in addition to resulting in a substantial increase in model fit relative to that of the CFA solution, it also resulted in well-defined factors. The parameter estimates from this bifactor-CFA solution are reported in the bottom section of Table 2 and reveal an adequately defined social behavior problems G-factor ($|\lambda|$=.01–.98; ω=.92) and strongly defined prosocial behavior S-factor ($|\lambda|$=.69–.81; ω=.91). In contrast, the aggressive behavior S-factor was more weakly defined ($|\lambda|$=.03–.68), although anchored in satisfactory factor loadings from three of its items, as illustrated by a ω value of .78. It is also noteworthy that the prosocial items tended to load more weakly on the social behavior problems G-factor ($|\lambda|$=.07–.31) relative to the aggressive items ($|\lambda|$=.28–.98). This solution was retained for further analyses, and proved to be fully invariant as a function of sex, age, ID level, and country (see Table S6 of the online supplements).

Concurrent Validity

The results from the predictive models estimated to assess the concurrent validity of youth, teacher, and parent reports of social behaviors are reported in Table 3.

Youth Reports of Social Behaviors. Focusing on youth self-reports of their social behaviors, the results show that the social behavior problems G-factor was associated with higher levels of youth, teacher, and parent ratings of peer victimization. The aggressive behaviors S-factor was associated with higher levels of youth ratings of peer victimization. The social behavior problems G-factor and the aggressive behaviors S-factor were both associated with higher levels of youth self-reports of depression. Finally, none of the youth-rated factors were associated with teacher and parent ratings of depression and hyperactivity-inattention. The prosocial behavior S-factor was not associated with the outcomes.

Teacher Reports of Youth's Social Behaviors. In relation to teacher reports of youth social behaviors, the results show that the social behavior problems G-factor was concurrently associated with higher levels on youth and teacher ratings of peer victimization and depression, as well as with higher levels on teacher ratings of hyperactivity-inattention. This social behavior problems G-factor was also linked to higher levels of parental ratings of peer victimization. Although associated with some youth-rated outcomes, the teacher-rated aggressive behavior S-factor was more consistently associated with teacher-rated outcomes (victimization, depression, and hyperactivity-inattention). Finally, the prosocial

behavior S-factor displayed a less conclusive pattern of associations with the outcomes. More precisely, this S-factor was found to be associated concurrently with slightly higher levels on teacher ratings of peer victimization, but not with other outcomes.

Parental Reports of Youth's Social Behaviors. The results from the analyses conducted using parental reports of youth social behaviors first show that these ratings were associated with parents' and teachers' ratings of depression and hyperactivity-inattention, as well as with youth and teacher ratings of peer victimization. More precisely, the results revealed that the social behavior problems G-factor was associated with higher levels on teacher and parental ratings of depression and hyperactivity-inattention, as well as with higher levels on youth and parental ratings of peer victimization. Similarly, the aggressive behavior S-factor was concurrently associated with higher levels on teacher and parental ratings of peer victimization and depression, as well as on parental ratings of hyperactivity-inattention. Finally, the prosocial behavior S-factor was concurrently associated with lower levels on teacher ratings of peer victimization, depression, and hyperactivity-inattention, as well as on parental ratings of depression.

Inter-Rater Agreement

We also assessed the level of inter-rater agreements in ratings of social behaviors by inspecting the correlations between raters reported in Table S2 of the online supplements. Interestingly, youth and teacher reports were not significantly correlated with one another, with the exception of youth's ratings on the prosocial behavior S-factor which were positively associated with teacher's ratings on the aggressive behavior S-factor ($r=.30$). Youth and parental reports were also uncorrelated, with the exception of their respective evaluation of the social behavior problems G-factor ($r=.30$). Teachers' and parents' evaluation of social behavior problems G-factor ($r=.27$), prosocial behavior S-factor ($r=.16$), and aggressive behavior S-factor ($r=.29$) were all significantly associated. All other associations were in the expected direction, with the exception of the association between teachers' ratings of the prosocial behavior S-factor and parents' ratings of the aggressive behavior S-factor ($r=.36$), which was positive.

Longitudinal Measurement Invariance and Test-Retest Stability

Results from the alternative measurement models estimated at T2 are reported in Tables 1 and 2, and essentially replicate the previously described T1 results regarding the superiority of adopting a bifactor-CFA solution for all three types of raters. Importantly, this bifactor-CFA proved to be fully invariant over time (see the bottom section of Table S3, S5, and S6 in the online supplements) for all three informants. For youth's self-reports, the results from the most invariant longitudinal model (i.e., latent mean invariance) revealed a one-year test-retest correlation of .58 for the social behavior problems G-factor, .32 for the prosocial behavior S-factor, and .54 for the aggressive behavior S-factor. For teachers, these results reveal a one-year test-retest correlation were of .77 for the social behavior problems G-factor, .50 for the prosocial behavior S-factor, and .93 for the aggressive behavior S-factor. Finally, for parents, the results reveal a one-year test-retest correlation of .78 for the social behavior problems G-factor, .26 for the prosocial behavior S-factor, and .54 for the aggressive behavior S-factor. It is important to note that these cannot be considered to provide a pure reflection of test-retest *reliability*, which is typically assessed over a much shorter time period (i.e., 1 week to a month) over which scores are expected to stay unchanged. In contrast, our results reflect test-retest *stability*, and thus the extent to which ratings can be expected to demonstrate some stability (encompassing a lack of random measurement error and a lack of true change).

Discussion

Calls have been made for the development of comprehensive measures of social behaviors for youth with ID (May & Kennedy, 2010). The present study addressed this call by proposing a new multi-informant (youth, parents, and teachers) measure of social behaviors designed for youth with ID and encompassing measures of prosocial and aggressive behaviors. This measure was simultaneously developed in English and French from items taken from the Strengths and Difficulties Questionnaire (Goodman, 2001) and the Children's Behavior Questionnaire (Tremblay et al., 1987), two well-established measures of social behaviors among TD youth which underwent substantial modifications to make them suitable for self-reports among youth with ID. Our results first showed that responses to this questionnaire provided by youth with ID, their teachers, and their parents followed a bifactor representation encompassing a global factor reflecting your global levels of social behavior problems (G-factor), along with two specific factors reflecting the extent to which the target youth displayed levels of prosocial and aggressive behaviors that deviated from that global social behavior problems G-

factor (the S-factors). Furthermore, our results also showed that each informant was able to provide a complementary perspective on the social behaviors of the target youth, consistent with the idea that youth tend to adjust their social behaviors to the unique characteristics of their environments, reinforcing the need to devise assessment procedures able to capture the unique perspective, voice, and agency of youth with ID.

More precisely, our results supported the psychometric properties of this multi-informant measure. Teacher and parental reports were found to work equivalently regardless of youth's sex, age, ID level, and country (corresponding to the linguistic version: English or French). As for youth self-reports, the measure performed equivalently between males and females, and across age categories, but evidenced some differential item functioning as a function of youth's ID level and linguistic version (see later discussion). Our results also supported the concurrent validity of the social behavior problems G-factor and of the aggressive behavior S-factor in relation to measures of victimization, depression, and hyperactivity-inattention both within and between informants. The concurrent validity of the prosocial behavior S-factors, however, was not as well-established, especially for youth and teacher reports, suggesting that these items should be mainly used to assess youth global levels of social behaviors rather than their specific levels of prosociality. Finally, we found that the psychometric properties of the youth, teacher, and parental versions of the instrument were replicated one year later and that each informants' perceptions displayed a moderate to high level of stability, especially for the social behavior problems G-factor and of the aggressive behavior S-factor.

Multiple Informants Provide Complementary Perspectives

A first conclusion was that all three informants were able to provide reliable, valid, and complementary perspectives on the social behaviors of the target youth. Indeed, although inter-rater reliability was low, composite reliability was high across raters, stability was moderately high across a period of one year, and the pattern of associations found in our analyses of convergent validity were similar in direction, number, variety, and intensity across all three informants. Our result thus reinforces the importance of being able to capture the unique perspective, voice, and agency of youth with ID which has been previously highlighted in research conducted among TD youth (Turk et al., 2012). When reporting on their own behaviors, youth's self-reports provide a unique opportunity to access characteristics of their social interactions occurring in contexts to which adult informants typically do not have access (e.g., recess, interactions with peers). For instance, some research has shown that school staff tend to underestimate students' aggressive behaviors (Bradshaw et al., 2007). This conclusion does not mean that informants' reports are irrelevant. Rather, it indicates that their value stems from their ability to uniquely complement youth self-reports (Prewett et al., 2019). Indeed, and in accordance with the idea that youth may adapt their behaviors to different life contexts (Aarts et al., 2003; Hubbard et al., 2010), parents and teachers are also able to provide a unique view of youth's behaviors as observed by adults within more structured (i.e., the classroom) or monitored (i.e., the household) contexts. Given that youth's, teacher's, and parent's perspectives capture youth's behaviors in only partially overlapping environments, it seems important that studies assessing such behaviors carefully choose the perspective that they want to reflect. We also emphasize that the unique perspectives of youth with ID are also crucial to measure as their perceptions are valid in their own right and also enable their voice and agency to be accounted for in research that affects them.

Differential Item Functioning and Discriminant Validity

Tests of measurement invariance revealed that teachers' and parents' reports functioned equally well irrespective of youth's characteristics (sex, age, ID level, and country/linguistic version), with no evidence of differential item functioning. However, youth's self-reports seemed to be slightly influenced by their ID level and by the linguistic version of the questionnaire (English or French). Importantly, evidence of differential item functioning was slightly more widespread in relation to youth's level of ID, encompassing the factor loadings associated with three items, and the response thresholds associated with three items (two of which had equal factor loadings). ID is known to affect cognitive processes (e.g., Visser et al., 2015) and representations of the self (Maïano et al., 2019) in a way that may explain these instances of differential item functioning. Despite this, our results suggest that the five items found to perform differently ("You forced others to give you something you wanted by scaring them"; "You have been attentive to other people's feelings"; You became physically aggressive or angry when someone hurt you"; "You became physically aggressive when something was taken away from you"; You have been helpful when someone was hurt or feeling unwell") might have

been too complex for youth with moderate levels of ID. Pending further research, these items should be used with caution in research seeking to compare youth with mild and moderate levels of ID. Minimally, latent variable methods should be used to control for their differential item functioning in analyses seeking to compare these two subpopulations of youth with ID. In contrast, differential item functioning between the English and French versions of the self-reported questionnaires was restricted to three response thresholds associated with a single item ("You have been helpful when someone was hurt or feeling unwell"). The fact that this item also performed differently across groups of youth with mild or moderate levels of ID clearly suggests that this item should be targeted for re-assessment, and perhaps reformulation. However, the fact that differential item functioning was limited to a single item does support the linguistic equivalence of the global measure, as long as proper caution is taken to handle this item properly (either by removing it from linguistic comparison, or controlling for this form of differential item functioning through latent variable methodologies).

In terms of discriminant validity, we found no differences between boys and girls, or between youth with mild and moderate ID levels. Among TD youth, girls tend to display more prosocial and fewer aggressive behaviors than boys (Hartl et al., 2020). Yet, in youth with ID, studies do not report such sex differences (Hove et al., 2010; Lakhan & Kishore, 2018), which is aligned with the lack of latent mean differences observed in the present studies in youth, teacher, and parental reports. Likewise, existing studies suggested a link between ID level and social behaviors (Lakhan & Kishore, 2018; Totsika et al., 2014), leading us to expect higher levels of aggressive behaviors and lower levels of prosocial behaviors in youth with moderate ID compared to those with mild ID. Yet, most of these previous studies also considered youth with severe and profound levels of ID, without clearly being able to establish differences between youth with mild and moderate levels of ID. Our results thus add to this body of research by suggesting that, pending replication, youth with mild and moderate levels of ID may not differ in terms of social behaviors.

Additionally, although we observed latent mean differences between Australian and Canadian youth, these differences were limited to youth self-reports. More precisely, Australian youth tended to report higher scores on the aggressive behavior S-factor (1.167 SD higher) relative to Canadian youth. Although this result needs to be interpreted while keeping in mind the bifactor nature of our instruments, and thus that this S-factor represents youth specific levels of aggressive behaviors as they deviate from their global levels of social behavior problems, this intriguing result clearly deserves further explanation. This result may reflect cultural or educational differences aligned with national stereotypes of Australians being more open, friendly, and outgoing – and thus less tolerant of aggressivity and violence – relative to Canadians, possibly leading Australian youth to be more self-critical regarding the possible aggressive nature of their behaviors. This explanation remains tentative, and future studies should more systematically explore the source of this difference.

Concurrent Validity

Tests of concurrent validity generally supported the concurrent validity of the social behavior problems G-factor and the aggressive behavior S-factor from the perspective of all three informants. More precisely, when youth, teachers, and parents perceive that the target youth displays a globally high level of social behavior problems, they also described them as being exposed to higher levels of victimization, and as presenting higher levels of depression and hyperactivity-inattention. This observation is consistent with results typically found by previous studies conducted among youth with ID (Clark et al., 2016; Davies & Oliver, 2014; Hartley & Birgenheir, 2009; Reiter & Lapidot-Lefler, 2007). Most of these associations were also found between informants, suggesting an excellent level of concurrent validity. Similarly, all three informants' ratings on the aggressive behavior S-factor were associated with higher levels of victimization, depression, and hyperactivity-inattention. However, parents' and teachers' perspectives seem to be more aligned with one another regarding the nature of the associations found between this dimension of social behaviors and youth's psychosocial functioning than with the associations reported by the target youth. Such results are also consistent with the finding that teachers' and parents' ratings of social behaviors were found to be more similar (correlated) to one another than to youth's self-reports.

However, the concurrent validity of the prosocial behavior S-factor was much weaker, and limited to parental reports. For instance, youth's ratings on the prosocial behavior S-factor were not associated with any of the outcomes, whereas teachers indicated that youth scoring higher on the prosocial behavior S-factor tended to present a higher risk of victimization. As this study was conducted

among adolescents, it is possible that students perceived by their teachers to be overly prosocial may actually be those described by their peers to be the "teachers' pet", and thus more likely to be picked on (e.g., Martin, 1984). However, the opposite argument is also plausible, suggesting that teachers may develop more positive perceptions of the students with whom they have to play a more protective role to defend them against peer victimization (Pozzoli & Gini, 2010). Some studies have found that teachers' ratings of the prosocial skills of students are more similar between victims and non-victims than ratings made by the students themselves or their peers (Fox & Boulton, 2005). In contrast, parental ratings of youth prosocial behaviors (S-factor) were those showing the best concurrent validity, especially in relation to teachers' and parents' ratings of depression and hyperactivity-inattention. Given that this S-factor captures youth prosocial behavior beyond their global level of social behavior problems, it is possible that parents, through their natural tendency to perceive their child positively (Brummeland et al., 2015), may be more sensitive to slight variations in prosocial behaviors, explaining why their ratings were the only one for which we found evidence of concurrent validity. Still, it is also important to note that other studies have also failed to establish concurrent validity of prosocial behavior with ratings of depression and hyperactivity-inattention among youth with ID (Rojahn et al., 2011). It is thus possible that the concurrent validity of this dimension would have been better established in relation to more positive (i.e., popularity, social integration, self-esteem) and/or with relational (i.e., teacher-student relationship quality) outcomes.

One-year Stability

This study provided replication evidence for the factor structure of our measure of social behaviors over one year. Responses obtained one year later by all informants matched the same factor structure, which was completely invariant over time. The results also revealed a moderate to high level of longitudinal stability within each informant. In this regard, our results revealed two noteworthy observations. First, and contrary to our expectations, teacher reports proved to be the most stable over time ($r=.50$ to .99), especially on the aggressive behavior S-factor, when compared to youth ($r=.32$ to .58) and parental reports ($r=.26$ to .78). This result is surprising considering that teachers change yearly, and might reflect the fact that teachers' ratings are circumscribed to a single environment, the classroom. As the classroom setting is a relatively uniform environment from one year to the next, youth might adopt similar behaviors over time (Aarts et al., 2003), explaining why different teachers share similar perceptions of youth's behaviors. In contrast, youth and parental ratings encompass several environments, such as the classroom and recess time for youth self-reports, and household and leisure activities for parental reports. Alternatively, these results could reflect the reputation that students come to develop in their schools, leading teachers to expect them to behave in a certain way. In a form of Pygmalion effect (Murdock-Perriera & Sedlacek, 2018), teachers' expectations could affect their judgment and actual student behavior.

Second, the prosocial behavior S-factor ($r=.26$ to .50) was the least stable component across all three informants relative to the social behavior problems G-factor ($r=.58$ to .78) and the aggressive behavior S-factor ($r=.54$ to .99). Beyond the relative stability of the social behavior problems G-factor, this result could indicate that youth's specific tendency to display kindness, altruism, and helpfulness may vary more importantly as a function of the context, consistent with some suggesting that prosocial behavior can result from spur of the moment environmentally influenced impulses (Baron, 1997). It is also consistent with studies showing that aggressive behaviors tend to result from more stable influences, whereas prosocial behaviors, although also relatively stable, are more likely to change depending on circumstantial influences (Lininga-Wijnen et al., 2018).

Limitations

The current study presents limitations. First, although we found that our results using a sample of youth with ID were similar to results obtained in studies using samples of TD youth, no comparison sample of TD youth was considered, thus making any claim of generalizability or difference tentative at best. Second, the present study relied on youth from two countries sharing a very similar culture and is thus unable to account for possible cultural differences. For instance, compared to Americans, Chinese youth have been found to place higher value on prosocial behaviors in their peer relationships and perceive aggression less favorably (Li et al., 2012). Future work is thus needed to verify the generalizability of our findings to more diversified samples of youth with ID, from a greater variety of countries and cultures. Third, while the current study allowed to verify the replicability of the factor structure over a one-year interval, tests of discriminant and convergent validity remained cross-sectional

in nature, and thus unable to inform questions related to the directionality of the observed associations. To better understand how social behaviors is predicted by and predicts psychosocial outcomes, future studies should rely on fully longitudinal research designs, making it possible to explicitly consider change, and the shape of change, in variables of interest as it occurs over time. Finally, although several precautions were taken to maximize and facilitate youth's understanding of the items, a few of them still selected the "do not understand" option. This observation highlights the importance for future studies to rely, as we did, on read-aloud assisted procedures and to incorporate formal or standardized comprehension checks when working with this population. Importantly, the use of the measure of social behaviors proposed in this article should, for the moment, be limited to populations of adolescents and young adults with mild to moderate levels of ID and adequate verbal (understanding) skills.

Conclusion

This study aimed to develop and validate a multi-informant scale capturing both the prosocial and aggressive aspects of the social behaviors of youth with ID. These two dimensions, although interconnected and studied as such among TD youth, have generally been studied independently among samples of youth with ID, possibly due to the lack of validated measurement instruments including both dimensions. Our results are encouraging regarding the ability of this new set of measures to accurately capture social behaviors among youth with mild to moderate levels of ID. Relying on a cross-cultural sample of Canadian and Australian youth with ID, the present study was also able to establish the adequacy of a French and English version of this measure, showing that it could be confidently used in both languages to assess social behaviors. This measure proved to be reliable and valid across the various verifications conducted as part of this study, and although the need to replicate the present results remains, this measure can now be confidently used to assess social behaviors among youth with ID. Importantly, our results also supported the idea that all three informants had a relevant and complementary perspective probably capturing the changing nature of youth behaviors. Despite these promising conclusions, some recommendations are in order. First, the results showed that caution is required when using youth self-reports to conduct comparisons between youth with mild and moderate levels of ID. Second, one item might be targeted for adaptation in future studies as it displayed evidence of differential item functioning as a function of youth ID level and linguistic version ("You have been helpful when someone was hurt or feeling unwell"). We hope that future research using tools like the one developed here will help researchers and practitioners construct and evaluate interventions to protect youth with ID from social behavior problems and their negative ramifications, as well as contribute to providing youth with ID with a voice and agency in a research context where current methodology privileges informant reports.

References

Aarts, H. & Dijksterhuis, A. (2003). The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology, 84*, 18-28.

APA (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th edition, text revised)*. American Psychological Association.

APA. (2020) *APA Dictionary of Psychology*. American Psychological Association.

Asparouhov, T., & Muthén, B.O. (2010). Weighted least square estimation with missing data. www.statmodel.com/download/GstrucMissingRevision.pdf

Asparouhov, T., Muthén, B.O., & Morin, A.J.S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561-1577.

Bailey, T., Totsika, V., Hastings, R.P., Hatton, C., & Emerson, E. (2019). Developmental trajectories of behaviour problems and prosocial behaviours of children with intellectual disabilities in a population-based cohort. *Journal of Child Psychology and Psychiatry*, *60*, 1210-1218.

Balboni, G., Mumbardó-Adam, C., Coscarelli, A. (2020). Influence of adaptive behaviour on the quality of life of adults with intellectual and developmental disabilities. *Journal of Applied Research in Intellectual Disability*, *33*, 584-594.

Baron, R. (1997). The sweet smell of helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. *Personality & Social Psychology Bulletin*, *23*, 498-503.

Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quinonez, H.R., & Young, S.L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 1-18.

Bradshaw, C.P., Sawyer, A.L., & O'Brennan, L.M. (2007) Bullying and peer victimization at school:

Perceptual differences between students and school staff. *School Psychology Review*, *36*, 361-382.

Brummelman, E., Thomaes, S., Nelemans, S. A., Orobio de Castro, B., & Bushman, B. J. (2015). My child is God's gift to humanity: Development and validation of the Parental Overvaluation Scale (POS). *Journal of Personality & Social Psychology, 108*, 665-679.

Capron, C., Thérond, C., & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment*, *23*, 79-88.

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Clark, M., Crocker, A.G., & Morin, D. (2016). Victimization history and aggressive behavior among adults with intellectual disabilities: The mediating role of mental health. *International Journal of Forensic Mental Health*, *15*, 301-311.

Craig, W.M., Pepler, D.J., & Atlas, R. (2000). Observations of bullying in the playground and in the classroom. *School Psychology International*, *21*(1), 22–36.

Cuthill, F.M., Espie, C.A., & Cooper, S. (2003). Development and psychometric properties of the Glasgow Depression Scale for people with a learning disability. *British Journal of Psychiatry*, *182*, 347-353.

Davies, L.E. & Oliver, C. (2014). The purported association between depression, aggression, and self-injury in people with intellectual disability: A critical review of the literature. *American Journal on Intellectual and Developmental Disabilities*, *119*, 452-471.

De Los Reyes, A. (2011). Introduction to the Special Section : More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*, 1-9.

Delforterie, M., Hesper, B. & Didden, R. (2018). Psychometric properties of the Dynamic Risk Outcome Scales (DROS) for individuals with mild intellectual disability or borderline intellectual functioning and externalizing behaviour problems. *Journal of Applied Research in Intellectual Disabilities*, *Special Issue 2018*, 1-11.

Dell'Armo, K.A. & Tassé, M.J. (2019). The role of adaptive behavior and parent expectations in predicting post-school outcomes for young adults with intellectual disability. *Journal of Autism and Developmental Disorders*, *49*, 1638-1651.

Enders, C.K. (2010). *Applied Missing Data Analysis*. Guildford.

Esbensen, A., Rojahn, J., Aman, M.G., & Ruedrich, S. (2003). Reliability and validity of an assessment instrument for anxiety, depression, and mood among individuals with mental retardation. *Journal of Autism and Developmental Disorders*, *33*, 617-629.

Finney, S.J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. Structural equation modeling: A second course* (pp. 439-492). IAP Information Age Publishing.

Fox, C.L. & Boulton, M.J. (2005). The social skills problems of victims of bullying: Self, peer, and teacher perceptions. *British Journal of Educational Psychology*, *75*, 313-328.

Furr, R.M., & Bacharach,V.R. (2014). *Psychometrics: An Introduction (2ⁿᵈ Ed)*. Thousand Oaks, CA: Sage.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1337-1345.

Hartl, A.C., Laursen, B., Cantin, S., & Vitaro, F. (2020). A test of the bistrategic control hypothesis of adolescent popularity. *Child Development*, *91*, e635-e648.

Hartley, S.L. & Birgenheir, D. (2009). Nonverbal social skills of adults with mild intellectual disability diagnosed with depression. *Journal of Intellectual Disability Research*, *2*, 11-28.

Hove, O. & Havik, O.E. (2008). Psychometric properties of Psychopathology checklists for Adults with Intellectual Disability (P-AID) on a community sample of adults with intellectual disability. *Research in Developmental Disabilities*, *29*, 467-482.

Hove, O. & Havik, O.E. (2010). Developmental level and other factors associated with symptoms of mental disorders and problem behaviour in adults with intellectual disabilities living in the community. *Social Psychiatry and Psychiatric Epidemiology*, *45*, 105-113.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Huang, C. (2017). Cross-informant agreement on the Child behavior Checklist for youths: A meta-analysis. *Psychological Reports*, *120*, 1096-1116.

Hubbard, J.A., McAuliffe, M.D., Morrow, M.T., & Romano, L.J. (2010). Reactive and proactive aggression in childhood and adolescence: Precursors, outcomes, processes, experiences, and measurement. *Journal of Personality*, *78*, 95-118.

Janosz, M., Bouthillier, C., Bowen, F., Chouinard, R., & Desbiens, N. (2007). *Rapport de validation du questionnaire sur l'environnement socioéducatif des écoles primaires*. Montréal, Qc: Université de Montréal.

Kurtek, P. (2018). Prosocial vs antisocial coping and general life satisfaction of youth with mild intellectual disability. *Journal of Intellectual Disability Research*, *62*, 581-592.

Lakhan, R. & Kishore, M.T. (2018). Behaviour problems in children with intellectual disabilities in a resource-poor setting in India – Part 1: Associations with age, sex, severity of intellectual disabilities and IQ. *Journal of Applied Research in Intellectual Disabilities*, *31*, 43-50.

Li, Y., Xie, H., & Shi, J. (2012). Chinese and American children's perceptions of popularity determinants: Cultural differences and behavioral correlates. *International Journal of Behavioral Development*, *36*, 420-429.

Lininga-Wijnen, L., Harakeh, Z., Dijkstra, J.K., Veenstra, R., & Vollebergh (2018). Aggressive and prosocial peer norms: Change, stability, and associations with adolescent aggressive and prosocial behavior development. *Journal of Early Adolescence*, *38*, 178-203.

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling*, *25*, 737-479.

Maïano, C., Bégarie, J., Morin, A.J.S., & Ninot, G. (2009). Assessment of physical self-concept in adolescents with intellectual disability: Content and factor validity of the very short form of the Physical Self-Inventory. *Journal of Autism and Developmental Disorders*, *39*, 775-787.

Maïano, C., Coutu, S., Morin, A.J.S., Tracey, D., Lepage, G., & Moullec, G. (2019). Self-concept research with school-aged youth with intellectual disabilities: A systematic review. *Journal of Applied Research in Intellectual Disabilities*, *32*, 238-255.

Maïano C., Morin A.J.S., Bégarie J., & Garbarino, J.-M. (2011). The Center for Epidemiologic Studies Depression Scale: Factor validity and reliability in a French sample of adolescents with mild to moderate intellectual disability. *Research in Developmental Disabilities*, *32*, 1872-1883.

Maïano C., Morin A.J.S., Bégarie J., & Ninot G. (2011). The Intellectual Disability Version of the Very Short Form of the Physical Self-Inventory (PSI-VS-ID): Cross-validation and measurement invariance across gender, weight, age and intellectual disability level. *Research in Developmental Disabilities*, *32*, 1652-1662.

Martin, W.B. (1984). Student perception of teachers' pets and class victims. *Canadian Journal of Education*, *9*, 89-99.

May, M.E. & Kennedy, C.H. (2010). Health and problem behavior among people with intellectual disabilities. *Behavior Analysis in Practice*, *3*, 4-12.

McDonald, R.P. (1970). Theoretical foundations of principal factor analysis and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1-21.

Méthot, S., & Morin, D. (2004). Traduction d'un instrument de dépistage des symptômes affectifs et anxieux chez les adultes ayant un retard mental. *Revue Francophone de la Déficience Intellectuelle*, *15*, 57-73.

Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.

Morin, A.J.S., Moullec, G., Maïano, C., Layet, L., Just, J.L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Epidemiology and Public Health*, *59*, 327-340.

Morin, A.J.S., Arens, A.K., & Marsh, H.W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, *23*, 116-139.

Morin, A.J.S., Myers, N.D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum & R.C. Eklund (Eds.), *Handbook of Sport Psychology*, 4th Edition (pp. 1044-1073). Wiley.

Murdock-Perriera, L.A. & Sedlacek, Q.C. (2018). Questioning Pygmalion in the twenty-first century:

The formation, transmission, and attributional influence of teacher expectancies. *Social Psychology of Education*, *21*, 691–707.

Muthén, L., & Muthén, B. (2019). *Mplus User's Guide*. Muthén & Muthén.

Oliver, P.C., Crawford, M.J., Rao, B., Reece, B., & Tyrer, P. (2007). Modified Overt Aggression Scale (MOAS) for people with intellectual disability and aggressive challenging behaviour: A reliability study. *Journal of Applied Research in Intellectual Disabilities*, *20*, 368-372.

Prewett, S.L., Bergin, D.A., & Huang, F.L. (2019). Student and teacher perceptions on student-teacher relationship quality. *School Psychology International*, *40*, 66-87.

Price, L.R. (2017). *Psychometric Methods: Theory and Practice*. New York: Guilford

Pozzoli, T. & Gini, G. (2010) Active defending and passive bystanding behavior in bullying: The role of personal characteristics and peer pressure. *Journal of Abnormal Child Psychology*, *38,* 815-827.

Quebec Longitudinal Study of Child Development. (2006, 2008). *Survey Description and methodology (Part 1)*. Institut de la Statistique du Québec.

Raykov, T., & Marcoulides, G.A. (2011). *Introduction to Psychometric Theory*. New York: Routledge

Reiter, S. & Lapidot-Lefler, N. (2007). Bullying among special education students with intellectual disabilities. *Intellectual and Developmental Disabilities*, *3*, 174-181.

Renk, K. (2005). Cross-informant ratings of the behavior of children and adolescents: The "Gold Standard." *Journal of Child and Family Studies*, *14*, 457-468.

Roid, G.H., & Miller, L.J. (1997). *Leiter International Performance Scale-Revised*. Stoelting.

Rojahn, J., Rowe, E.W., Kasdan, S. Moore, L. & van Ingen, D.J. (2011). Psychometric properties of the Aberrant Behavior Checklist, the Anxiety, Depression and Mood Scale, the Assessment of Dual Diagnosis and the Social Performance Survey Schedule in adults with intellectual disabilities. *Research in Developmental Disabilities*, *32*, 2309-2320.

Rojahn, J., Rowe, E.W., Sharber, A.C., Hastings, R., Matson, J.L., Didden, R., Kroes, D.B.H., & Dumont, E.L.M. (2012a). The Behavior Problems Inventory Short Form for individuals with intellectual disabilities: Part I: Development and provisional clinical reference data. *Journal of Intellectual Disability Research*, *56*, 527–545.

Rojahn, J., Rowe, E.W., Sharber, A.C., Hastings, R., Matson, J.L., Didden, R., Kroes, D.B.H., & Dumont, E.L.M. (2012a). The Behavior Problems Inventory-Short Form for individuals with intellectual disabilities: Part II: Reliability and validity. *Journal of Intellectual Disability Research*, *56*, 546–565.

Salmivalli, C. & Nieminen, E. (2002). Proactive and reactive aggression in bullies, victims, and bully–victims. *Aggressive Behavior*, *28*, 30–44.

Sparrow, S.S., Cicchetti, D.V., Saulnier, C.A. (2016). *Vineland Adaptive Behavior Scales, Third Edition*. Circle Pines.

Tassé, M.J., Balboni, G., Navas, P., Luckasson, R., Nygren, M.A., Belacchi, C., Bonichini, S., Reed, G.M., & Kogan, C.S. (2019) Developing behavioural indicators for intellectual functioning and adaptive behaviour for ICD-11 disorders of intellectual development. *Journal of Intellectual Disability Research*, *63*, 386-407.

Tremblay, R.E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The preschool behaviour questionnaire : Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development*, *10*, 467-484.

Tsiouri, J.A., Kim, S.Y., Brown, W.T., & Cohen, I.L. (2011). Association of aggressive behaviours with psychiatric disorders, age, sex and degree of intellectual disability: A large-scale survey. *Journal of Intellectual Disability Research*, *55*, 636-649.

Totsika, V., Hastings, R.P., Vagenas, D., & Emerson, E. (2014). Parenting and the behavior problems of young children with an intellectual disability: Concurrent and longitudinal relations in a population-based study. *American Journal on Intellectual & Developmental Disabilities*, *119*, 422-435.

Turk, V., Khattran, S., Kerry, S., Corney, R., & Painter, K. (2012). Reporting of health problems and pain by adults with an intellectual disability and by their carers. *Journal of Applied Research in Intellectual Disabilities*, *25*, 155-165.

van den Bogaard, K.J.H.M., Nijman, H.L.I., Palmstierna, T. & Embregts, P.J.C.M. (2018). Characteristics of aggressive behavior in people with mild to borderline intellectual disability and co-occuring psychopathology. *Journal of Mental Health Research in Intellectual Disabilities*, *11*, 124-142.

Visser, E.M., Berger, H.J.C., Van Schrojenstein Lantman-De Valk, H.M.J, Prins, J.B., & Teunisse, J.P. (2015). Cognitive shifting and externalising problem behaviour in intellectual disability and autism spectrum disorder. *Journal of Intellectual Disability Research*, *59*, 755-766.

Weschler, D. (2008). *Weschler Intelligence Scale for Children-Fourth Edition*. Pearson.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California.

Table 1

Measurement Models for Youth, Teacher, and Parental Reports of Youth Social behaviors.

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI
Youth Self-Reports T1						
1. One-factor CFA	507.11*	35	.84	.79	.20	.18; .21
2. Two-factor CFA	71.26*	34	.99	.98	.06	.04; .07
3. Two-factor ESEM	55.77*	26	.99	.98	.06	.04; .08
4. Bifactor-CFA	34.65	25	1.00	1.00	.03	.00; .05
5. Bifactor-ESEM	20.15	18	1.00	1.00	.02	.00; .05
Youth Self-Reports T2						
1. One-factor CFA	352.70*	35	.80	.74	.19	.18; .21
2. Two-factor CFA	41.92	34	1.00	.99	.03	.00; .06
3. Two-factor ESEM	41.64*	26	.99	.98	.05	.02; .08
4. Bifactor-CFA	27.85	25	1.00	1.00	.02	.00; .06
5. Bifactor-ESEM	19.83	18	1.00	1.00	.02	.00; .06
Teacher Reports T1						
1. One-factor CFA	852.70*	90	.88	.86	.17	.16; .18
2. Two-factor CFA	340.76*	89	.96	.95	.10	.09; .11
3. Two-factor ESEM	374.58*	76	.98	.94	.12	.11; .13
4. Bifactor-CFA	268.84*	75	.97	.96	.10	.08; .11
5. Bifactor-ESEM	201.34*	63	.98	.96	.09	.08; .10
Teacher Reports T2						
1. One-factor CFA	1848.33*	90	.83	.80	.37	.35; .38
2. Two-factor CFA	214.76*	89	.99	.99	.10	.08; .12
3. Two-factor ESEM	187.59*	76	.99	.99	.10	.08; .12
4. Bifactor-CFA	152.42*	75	.99	.99	.08	.07; .10
5. Bifactor-ESEM	122.08*	63	.99	.99	.08	.06; .10
Parental Reports T1						
1. One-factor CFA	1873.44*	90	.54	.46	.33	.32; .35
2. Two-factor CFA	183.63*	89	.98	.97	.08	.06; .09
3. Two-factor ESEM	144.72*	76	.98	.98	.07	.05; .09
4. Bifactor-CFA	120.56*	75	.99	.98	.06	.04; .08
5. Bifactor-ESEM	70.77	63	1.00	1.00	.03	.00; .05
Parental Reports T2						
1. One-factor CFA	477.12*	90	.78	.75	.20	.18; .21
2. Two-factor CFA	133.28*	89	.98	.97	.07	.04; .09
3. Two-factor ESEM	158.99*	76	.95	.94	.10	.08; .12
4. Bifactor-CFA	99.24*	75	.99	.98	.05	.02; .08
5. Bifactor-ESEM	86.94*	63	.99	.98	.06	.02; .09

Note. *p < .01; χ^2: Chi square test of model fit and associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker–Lewis Index; RMSEA: Root Mean Square Error of Approximation and 90% Confidence Interval (CI); Δ: Change according to the previous retained model; Δχ^2: Chi square difference test calculated using the Mplus DIFFTEST option.

Table 2

Standardized Factor Loadings (λ) and Uniquenesses (δ) from the Final Measurement Models

Items	G-Factor (Socio. Beh. Prob.)		S-Factor (Prosocial)		S-Factor (Aggressive)			
	λ_{T1}	λ_{T2}	λ_{T1}	λ_{T2}	λ_{T1}	λ_{T2}	δ_{T1}	δ_{T2}
Youth Reports								
Considerate	-.64	-.27	*.21*	.64			.55	.52
Share	-.54	-.28	.46	.55			.50	.62
Helpful when someone hurts	-.79	-.39	.32	.75			.27	.29
Play with friends	-.19	*-.17*	.74	.58			.42	.64
Help others	-.56	-.18	.61	.73			.32	.43
Aggressive when teased	.47	.79			.77	*.32*	.18	.27
Aggressive when contradicted	.33	.93			.83	*.21*	.20	.10
Forced others to do things	.47	.59			.52	.47	.51	.43
Aggressive when hurt or angry	.44	.61			.74	.51	.26	.37
Aggressive if something taken	.43	.59			.70	.76	.33	.08
Composite reliability (ω)	.87	.86	.73	.81	.90	.80		
Scale score reliability (α)	.81	.78	.79	.77	.85	.86		
α for youth with Mild ID	.81	.73	.77	.77	.86	.84		
α for youth with Moderate ID	.80	.80	.80	.78	.83	.86		
Teacher Reports								
Considerate	-.73	-.47	.40	.67			.31	.33
Shares	-.63	-.56	.48	.68			.38	.22
Helpful	-.57	-.53	.80	.75			*.04*	.15
Kind to younger	-.81	-.66	.21	.39			.49	.44
Offers to help	-.57	-.50	.39	.51			*.05*	.01
One good friend	-.70	-.59	*.12*	.46			.30	.42
Liked by others	-.96	-.99	-.18	*-.14*			.52	.49
Aggressive when teased	.45	.52			.73	.82	.26	.06
Aggressive when contradicted	.49	.55			.76	.82	.18	.03
Violent when hurt	.47	.45			.67	.75	.25	.20
Said bad things	.35	.37			.63	.75	.33	.24
Aggressive if something taken	.52	.59			.75	.80	.48	.30
Scared others to get something	.48	.59			.73	.67	.13	.14
Physically attacked others	.25	.58			.78	.72	.17	.02
Hit, bitten, kicks others	.50	.55			.74	.73	.20	.16
Composite reliability (ω)	.95	.96	.76	.87	.94	.97		
Scale score reliability (α)	.67	.75	.89	.88	.91	.94		
Parent Reports								
Considerate	-.25	-.61	.69	*.38*			.46	.49
Shares	-.23	-.81	.70	*.18*			.45	.32
Helpful	-.01	-.54	.81	.84			.35	.00
Kind to younger	-.29	-.81	.79	*-.12*			.42	.52
Offers to help	-.31	-.56	.72	*.20*			.41	.41
One good friend	-.07	-.69	.76	*-.06*			.29	.33
Liked by others	-.19	-.72	.74	*-.27*			.38	.65
Aggressive when teased	.93	*.18*			*.03*	.88	.14	.19
Aggressive when contradicted	.98	.20			*.09*	.92	*.04*	.12
Violent when hurt	.63	*.15*			.62	.76	.36	.19
Said bad things	*.28*	.00			.68	.55	.23	.39
Aggressive if wants something	.81	.26			.22	.86	.47	.69
Scared others to get something	.66	*.19*			.45	.88	.15	.22
Physically attacked others	.86	.37			*.33*	.80	.29	.19
Hit, bitten, kicks others	.85	.42			*.30*	.79	.18	.20
Composite reliability (ω)	.92	.90	.91	.60	.80	.95		
Scale score reliability (α)	.77	.74	.88	.84	.89	.88		

Note. Non-statistically significant results (p ≤ .05) are marked in italics; G: Global factor from a bifactor solution; S: Specific factor from a bifactor solution; ω: omega coefficient of composite reliability (McDonald, 1970); α: Cronbach's alpha.

Table 3

Concurrent Validity.

	Victimization (SR)			Victimization (TR)			Victimization (PR)			Depression (SR)			Depression (TR)			Depression (PR)			H/I (TR)			H/I (PR)		
	b	s.e.	β	b	s.e.	β	b	s.e.	β	b	s.e.	β	b	s.e.	β	b	s.e.	β	b	s.e.	β	b	s.e.	β
Youth Reports																								
Age	-.03	.03	-.06	-.03	.01**	-.15	.01	.02	.05	-.02	.02	-.05	.00	.02	.02	-.02	.02	-.09	-.03	.03	-.08	-.02	.03	-.07
Sex (0=girl)	.05	.11	-.02	.00	.05	.00	-.01	.07	-.01	-.18	.08*	-.11	-.04	.06	-.04	-.03	.07	-.03	.09	.10	.06	.07	.11	.05
ID (0=mild)	.27	.10**	.11	-.11	.05*	-.15	.04	.07	.05	-.14	.08	-.09	-.03	.06	-.03	-.06	.08	-.06	.20	.10*	.13	.13	.11	.09
G-SBP	**.56**	**.08****	**.37**	**.08**	**.03***	**.15**	**.14**	**.06***	**.23**	**.35**	**.05****	**.35**	.08	.05	.14	.07	.06	.12	.09	.08	.10	-.10	.08	-.11
S-Prosocial	-.07	.08	-.04	-.03	.03	-.05	.05	.05	.07	-.03	.07	-.03	.05	.05	.08	.01	.06	.01	-.01	.09	-.01	-.03	.08	-.03
S-Aggressive	**.69**	**.08****	**.43**	**.08**	**.04***	**.16**	-.00	.05	-.01	**.19**	**.06****	**.17**	.06	.05	.10	.04	.05	.06	.08	.09	.07	-.12	.08	-.12
Teachers' Reports																								
Age	-.07	.04	-.05	-.01	.01	-.06	.02	.02	.11	-.01	.02	-.01	.03	.01*	.11	-.02	.02	-.07	.00	.02	.00	-.03	.03	-.07
Sex (0=girl)	.01	.13	.00	-.02	.05	-.03	-.02	.07	-.02	-.18	.09*	-.11	-.08	.05	-.09	-.04	.07	-.04	.02	.09	.01	.04	.11	.03
ID (0=mild)	.17	.13	.07	-.10	.04*	-.14	.04	.07	.05	-.18	.09*	-.11	-.01	.05	-.01	-.07	.07	-.07	.22	.08**	.15	.12	.11	.09
G-SBP	**.28**	**.10****	**.22**	**.14**	**.03****	**.32**	**.13**	**.05***	**.25**	**.15**	**.06***	**.18**	**.27**	**.03****	**.52**	.08	.06	.14	**.28**	**.05****	**.33**	.07	.08	.08
S-Prosocial	.09	.12	.06	**.06**	**.02****	**.13**	.06	.05	.11	.02	.07	.02	.04	.03	.07	-.02	.05	-.03	-.03	.06	-.04	-.08	.10	-.08
S-Aggressive	**.23**	**.10***	**.15**	**.19**	**.03****	**.38**	.02	.07	.03	.12	.08	.11	**.16**	**.03****	**.26**	.07	.08	.12	**.36**	**.07****	**.36**	.03	.09	.04
Parental Reports																								
Age	-.03	.04	-.05	-.03	.01*	-.14	.04	.01*	.15	-.01	.02	-.03	.01	.02	.04	.00	.02	.01	-.02	.02	-.04	-.00	.02	-.01
Sex (0=girl)	.02	.14	.06	.01	.05	.01	-.02	.07	-.02	-.15	.09	-.10	-.05	.06	-.05	-.04	.06	-.04	.06	.10	.04	.03	.10	.02
ID (0=mild)	.13	.13	.01	-.13	.05**	-.17	-.10	.06	-.11	-.20	.09*	-.13	-.07	.06	-.07	-.20	.07**	-.20	.13	.10	.09	-.02	.10	-.01
G-SBP	**.33**	**.14***	**.22**	.05	.05	.10	**.27**	**.06****	**.44**	.05	.08	.05	**.14**	**.06***	**.22**	**.25**	**.05****	**.38**	**.24**	**.10***	**.24**	**.28**	**.08****	**.29**
S-Prosocial	-.06	.13	-.05	**-.07**	**.03***	**-.16**	.03	.03	.06	-.07	.06	-.09	**-.09**	**.05***	**-.17**	**-.08**	**.02****	**-.16**	**-.19**	**.08***	**-.23**	-.05	.05	-.06
S-Aggressive	-.22	.13	-.12	**.11**	**.06***	**.18**	**.20**	**.05****	**.27**	.03	.10	.02	**.15**	**.07***	**.21**	**.23**	**.06****	**.30**	.16	.13	.14	**.23**	**.09****	**.21**

Note. *p* < .05; **p* < .01; Statistically significant coefficients for validity analyses are marked in bold; Within-informant paths are framed; b: Unstandardized regression coefficient; s.e. Standard error of the coefficient; β: Standardized regression coefficient; ID: Intellectual disability; G-SBP: Social behavior problems global factor; S-Prosocial: Prosocial behavior specific factor; S-Aggressive: Aggressive behavior specific factor; SR: Youth reports; TR: Teacher reports; PR: Parental reports; H/I: Hyperactivity/Inattention; Victim: Peer victimization; Depress: Depression.

Online Supplements for

Development and Validation of a Multi-Informant Measure of Social Behaviors for Youth with

Intellectual Disabilities

<div align="center">**Scale Development: Pilot Procedures**</div>

Objectives

The first objective of these pilot procedures were to examine the appropriateness of the format and clarity of the social behavior items (i.e., prosocial and aggressive behaviors) drawn from the Strengths and Difficulties Questionnaire (Goodman, 2001; Capron et al., 2007) and of the deviant behavior items drawn from the Children's Behavior Questionnaire (Quebec Longitudinal Study of Child Development, 2006, 2008; Tremblay et al., 1987) for use with youth with ID. Importantly, both of the questionnaires from which these items were taken were already available in French and English. Following this initial verification, the items were adapted to increase their clarity and ease of application based on recommendations related to the use of self-report questionnaires among people with ID (Finlay & Lyons, 2001, 2002). This preliminary adaptation was then tested among a first sample of youth with ID, which lead to further adaptations. The final adaptation was tested again among a second sample of youth with ID.

Method

Participants and Procedures. The pilot sample included 34 youth (aged between 13 and 21 years; 35% girls) with mild to moderate-severe ID, including 20 English-speaking Australians and 14 French-speaking Canadians. A first subsample of 18 youth (N = 10 in Australia and 8 in Canada) was solicited to evaluate the format and clarity of a preliminary adaptation of our measures. A second subsample of 16 youth (N = 10 in Australia and 6 in Canada) was solicited to assess the format and clarity of the final adapted version of our questionnaire. The procedures used in this pilot study were identical to those used in the main study, and received approval from the same research ethics committees. However, in the pilot process, the social behavior items were administered individually, at school, by a trained research assistant using a read-aloud assisted procedure to maximize youth's understanding and to facilitate discussion. The administration was mainly focused on assessing the level of understanding of the youth and the ease with which they could respond to the items.

Measures. A preliminary assessment of the appropriateness of the format and clarity of the items was conducted by all members of the research team familiar with the use of self-report questionnaires among youth with ID. This preliminary assessment revealed that the item format was potentially problematic for use as a self-reported questionnaire (rather than as an individually administered questionnaire) among youth with ID. A first concern was related to the interrogative format of the questions (e.g., *Have you shared your belongings with others?*), which could potentially inflate youth's tendencies to respond in an extreme either-or manner (i.e., Yes or No), which is reinforced by the tendency of youth with ID to acquiesce when responding to questions (Finlay & Lyons, 2002), rather than as a matter of degree. For this reason, the questions were reformulated in an affirmative manner (e.g., *You shared your belongings with others.*) and maximally simplified. To further increase youth's understanding of the sentences, words from the items were also associated with pictograms (presented above the words). The response scale used for all items was also converted to a five-point response scale. Indeed, although both original questionnaires relied on a three-point response scale (e.g., not true, somewhat true, and certainly true), we felt that this type of response scale lacked precision, and was too different from the response scale used on the other measures administered in this study. We thus extended this response scale to a six point (0 to 5) frequency scale, which was defined using pictograms reflecting the frequencies to which each behavior has occurred over the past week using a hand showing no fingers to a hand showing five fingers. Additionally, a "do not understand the statement" option was added to the response scale for situations in which respondents remained unable to understand the item. During this process, decisions were taken by consensus among research team members, as well as through consultation with school personnel (i.e., teachers, psychologists, and psycho-educators) familiar with youth with ID. The resulting version of our questionnaire, following this initial process of adaptation, was administered to the first pilot sample of students.

Results

The responses provided by the first subsample of youth revealed that some words used in some of the items were hard to understand for youth with ID (more specifically by those with more severe levels of ID). These results also revealed that the adjusted response scale seemed easy to understand for all participants, but that some participants still tended to respond to the items via a simpler "yes" or "no". Therefore, to further increase youth's understanding of the graphical six-point response scale, words were added to the pictograms (i.e., the hand showing no to the hand showing five fingers): Never,

1 time, 2 times, 3 times, four times, and 5 times or more. In addition, these verbal anchors were accompanied by the incorporation of a "no" or "yes" corresponding to youth more natural tendencies placed above the response scale to indicate that "yes" answered needed to be expanded for precision (i.e., "never" was associated with "no" and "1 time" to "5 times or more" were associated with "yes").

Finally, a template comprising a graphical displays and pictograms was developed to explain to youth how to use the response scale. This revised version was administered to the second subsample of youth. Results supported the adequacy of the final French and English adapted versions of the social behavior items and proved their suitability for use as self-report instruments among youth with ID.

References

Capron, C., Thérond, C., & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment*, *23*, 79-88.

Finlay, W.M.L., & Lyons, E. (2001). Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. *Psychological Assessment*, *13*, 319–335.

Finlay, W.M.L., & Lyons, E. (2002). Acquiescence in interviews with people with mental retardation. *Mental Retardation*, *40*, 14–29.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1337-1345.

Quebec Longitudinal Study of Child Development. (2006, 2008). *Survey Description and methodology (Part 1)*. Institut de la Statistique du Québec.

Tremblay, R.E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The preschool behaviour questionnaire : Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development*, *10*, 467-484.

Table S1

Use of the Do Not Understand Response Option and Missing Responses

	DNU				Missing (including DNU)			
	%	Correlation (*r*) with Global IQ	Correlation (*r*) with Verbal IQ	Mild-Moderate Difference (*p*)	%	Correlation (*r*) with Global IQ	Correlation (*r*) with Verbal IQ	Mild-Moderate Difference (*p*)
Social skills items (5)								
Attentive to others' feelings	2.30%	-.09	-.14	.32	3.45%	-.08	.01	.91
Share	.29%	(empty cells)	(empty cells)	.95	1.73%	-.09	(empty cell)	.37
Helpful when someone hurts	.57%	-.09	-.13	.18	2.01%	-.16*	-.13	.05[c]
Play with friends	.29%	(empty cells)	(empty cells)	.19	1.44%	-.02	.21*	.71
Help others	1.15%	-.16*	-.13	.18	2.30%	-.19**	-.13	.01[d]
Aggression items (5)								
Aggressive when teased	.57%	.03	.02	.17	1.43%	.03	.02	.71
Aggressive when contradicted	2.30%	-.08	.08	.94	3.16%	-.02	.08	.84
Forced others to do things	1.72%	-.18**	(empty cells)	.01[a]	2.29%	-.16*	.05	.14
Aggressive when hurt or angry	1.15%	-.11	(empty cells)	.04[b]	2.01%	-.13	.05	.23
Aggressive if something taken	.57%	-.03	(empty cells)	.96	2.01%	-.11	-.07	.64

Note. %DNU: Proportion of youth who selected the do not understand option for each items. *r*: Pearson's correlation. *p*: p-value. a: Mild ID: 0% DNU; Moderate ID: 4.03% DNU. b: Mild ID: 0% DNU; Moderate ID: 2.67% DNU; c: Mild ID: 0.61% Missing; Moderate ID: 3.87% Missing. d: Mild ID: 0% Missing; Moderate ID: 4.52% Missing.

Table S2

Correlations and Descriptive Statistics

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1.	Age														
2.	Sex (0=girl)	-.09													
3.	ID level (0=mild)	-.23**	.01												
4.	G-SBP (SR) T1	.15*	.00	.01											
5.	S-Prosocial (SR) T1	.19*	.07	-.01	.00										
6.	S-Aggressive (SR) T1	-.28**	.11	-.16*	.00	.00									
7.	G-SBP (TR) T1	-.11	.10	.02	.13	-.08	.16								
8.	S-Prosocial (TR) T1	-.01	-.06	.02	.06	.04	.04	.00							
9.	S-Aggressive (TR) T1	-.20**	.05	-.13	.18	.30**	.17	.00	.00						
10.	G-SBP (PR) T1	-.29**	.24**	.17*	.30**	.18	.10	.27**	.07	.48**					
11.	S-Prosocial (PR) T1	.12	-.19	.17	.03	.22	-.12	-.30**	.16*	-.09	.00				
12.	S-Aggressive (PR) T1	.01	-.09	-.08	.09	-.03	-.13	.27*	.36**	.29*	.00	.00			
13.	Victim. (SR) T1	-.09	.04	.04	.44**	-.10	.44**	.18**	.03	.20**	.32**	-.12	-.05		
14.	Victim. (TR) T1	.08	-.01	.02	.12	-.07	.23**	.29**	.20**	.38**	.10	.24	-.15	.15	
15.	Victim. (PR) T1	-.18**	.04	.21**	.29*	.09	-.03	.22*	.10	.10	.33**	.35**	.06	.27**	.25**
16.	Depression (SR) T1	-.08	-.08	.21**	.43**	-.03	.23**	.19**	.03	.17*	.10	.02	-.10	.43**	.08
17.	Depression (TR) T1	-.06	-.03	.12*	.23**	.14	.11	.48**	.00	.27**	.19*	.34**	-.17*	-.06	.46**
18.	Depression (PR) T1	-.03	-.06	.11	.10	-.03	.10	.17	.05	.15	.33**	.37**	-.15*	.21**	.18*
19.	H/I (TR) T1	-.12	.07	-.20**	.13	.00	.12	.36**	-.01	.40**	.30**	.15	-.31**	.16**	.09
20.	H/I (PR) T1	-.06	.05	.01	-.13	-.10	-.15	.15	-.02	.03	.25**	.29**	-.10	-.17*	.27**
21.	G-SBP (SR) T2	-.15	-.07	-.06	.42**	.21	.33**	.30**	-.09	.31*	.28*	-.10	.360*	.36**	.15
22.	S-Prosocial (SR) T2	-.04	.07	-.03	-.21*	.27*	-.03	-.13	.17	.12	.09	.36**	-.05	-.10	-.07
23.	S-Aggressive (SR) T2	-.20	.18	-.02	.13	-.26	.37**	.35*	.02	-.25	-.16	-.22	-.55*	.24**	.08
24.	G-SBP (TR) T2	-.11	.03	.12	.15	-.18	.27*	.66**	-.02	.32**	.52**	-.34**	.17	.22*	.45**
25.	S-Prosocial (TR) T2	.02	-.12	-.12	.05	.24	-.26	-.24**	.35**	.10	-.02	.30**	.19	-.10	.18
26.	S-Aggressive (TR) T2	-.14	.05	-.04	.39**	.14	.07	.13	.08	.52**	.31*	-.32**	.13	.22**	.42**
27.	G-SBP (PR) T2	-.24*	.09	.01	-.12	.06	.18	.45**	.02	.19	.19	-.73**	.00	.06	.21
28.	S-Prosocial (PR) T2	-.34**	.21*	.07	.15	-.11	-.35*	.08	.20	.16	-.02	-.16	.22	.06	.04
29.	S-Aggressive (PR) T2	.03	.22*	-.14	.28*	-.24	.17	-.07	.00	.50**	.67**	-.19	.47**	-.18	.11
Mean		15.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.97	1.27
S.D.		2.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.18	.38

Table S2 (continued)

Correlations and Descriptive Statistics

	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.
15. Victim. (PR) T1															
16. Depression (SR) T1	.17**														
17. Depression (TR) T1	.06	.08													
18. Depression (PR) T1	.39**	.30**	.13												
19. H/I (TR) T1	.31**	.21**	.15	.50**											
20. H/I (PR) T1	.02	-.14	.56**	-.02	.32**										
21. G-BAP (SR) T2	.31**	.36**	.28**	.28*	.17	-.08									
22. S-Prosocial (SR) T2	-.06	-.11	.11	.03	.09	.10	.00								
23. S-Aggressive (SR) T2	-.05	.15	-.08	-.36*	.27*	-.27*	.00	.00							
24. G-SBP (TR) T2	.31**	.05	.46**	.27**	.47**	.21*	.31**	-.07	-.06						
25. S-Prosocial (TR) T2	.05	.15	-.21*	-.05	-.07	.07	-.27*	.19	-.20	.00					
26. S-Aggressive (TR) T2	.34**	.11	.24**	.24*	.23*	.09	.27	.26	.12	.00	.00				
27. G-SBP (PR) T2	.11	.25*	.21	.28*	.31**	.07	.50**	-.37**	.29*	.40**	-.35*	.31*			
28. S-Prosocial (PR) T2	.18	-.06	.12	.06	-.07	.12	.54**	.00	-.29	.44**	.15	.51**	.00		
29. S-Aggressive (PR) T2	.16	.05	.07	.22*	.18	.04	-.15	.04	.19	-.05	.24	.19	.00	.00	
Mean	1.37	1.71	1.70	1.70	1.97	1.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S.D.	.46	.77	.48	.48	.79	.72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note. *p* < .05; **p* < .01; all latent factors are estimated with a mean of 0 and a standard deviation of 1. G-SBP: Social behavior problems global factor; S-Prosocial: Prosocial behavior specific factor; S-Aggressive: Aggressive behavior specific factor; SR: Youth self-report; TR: Teacher report; PR: Parental report; Victim.: Victimization; H/I: Hyperactivity-inattention.

Table S3

Measurement Invariance Models for Youth Self-Reported Social Behaviors.

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA
Measurement Invariance (Bifactor-CFA): Sex											
1. Configural invariance	59.42	50	1.00	.99	.03	.00; .06	–	–	–	–	–
2. Weak invariance	73.07	67	1.00	1.00	.02	.00; .05	17.46	17	.00	+.01	-.01
3. Strong invariance	103.01	100	1.00	1.00	.01	.00; .04	31.61	33	.00	.00	-.01
4. Strict invariance	113.32	110	1.00	1.00	.01	.00; .04	11.64	10	.00	.00	.00
5. Latent variance-covariance invariance	114.56	113	1.00	1.00	.01	.00; .04	2.73	3	.00	.00	.00
6. Latent mean invariance	117.77	116	1.00	1.00	.01	.00; .04	3.16	3	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Age[1]											
1. Configural invariance	74.64*	50	.99	.98	.06	.03; .08	–	–	–	–	–
2. Weak invariance	91.48*	67	.99	.99	.05	.02; .07	23.45	17	.00	+.01	-.01
3. Strong invariance	125.72*	100	.99	.99	.04	.01; .06	35.45	33	.00	.00	-.01
4. Strict invariance	136.78*	110	.99	.99	.04	.01; .06	11.00	10	.00	.00	.00
5. Latent variance-covariance invariance	124.54	113	1.00	1.00	.03	.00; .05	1.18	3	+.01	+.01	-.01
6. Latent mean invariance	134.80	116	.99	.99	.03	.00; .06	5.58	3	-.01	-.01	.00
Measurement Invariance (Bifactor-CFA): ID Level[1]											
1. Configural invariance	47.90	50	1.00	1.00	.00	.00; .05	–	–	–	–	–
2. Weak invariance	84.64	67	.99	.99	.04	.00; .07	32.01	17	-.01	-.01	+.04
2. Weak invariance–partial	62.08	64	1.00	1.00	.00	.00; .05	14.51	14	.00	.00	.00
3. Strong invariance	114.90	97	.99	.99	.03	.00; .06	54.26	33	-.01	-.01	+.03
3. Strong invariance–partial	97.21	94	1.00	1.00	.02	.00; .05	36.71	30	.00	.00	+.02
4. Strict invariance	113.44	104	1.00	1.00	.02	.00; .05	17.00	10	.00	.00	.00
5. Latent variance-covariance invariance	125.99	107	.99	.99	.03	.00; .06	6.67	3	-.01	-.01	+.01
6. Latent mean invariance	115.02	110	1.00	1.00	.02	.00; .04	.82	3	+.01	+.01	-.01
Measurement Invariance (Bifactor-CFA): Country											
1. Configural invariance	55.38	50	1.00	1.00	.03	.00; .06	–	–	–	–	–
2. Weak invariance	65.87	67	1.00	1.00	.00	.00; .04	13.78	17	.00	.00	-.03
3. Strong invariance	111.87	100	1.00	1.00	.03	.00; .05	47.91	33	.00	.00	+.03
3. Strong invariance–partial	98.70	97	1.00	1.00	.01	.00; .04	32.62	30	.00	.00	+.01
4. Strict invariance	105.12	107	1.00	1.00	.00	.00; .04	6.92	10	.00	.00	-.01
5. Latent variance-covariance invariance	104.84	110	1.00	1.00	.00	.00; .03	1.88	3	.00	.00	.00
6. Latent mean invariance	278.15*	113	.94	.95	.09	.08; .11	47.88*	3	-.06	-.05	+.09
7. Latent mean invariance–partial	110.46	112	1.00	1.00	.00	.00; .04	2.67	2	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Time											
1. Configural invariance	127.77	131	1.00	1.00	.00	.00; .02	–	–	–	–	–
2. Weak invariance	153.02	148	1.00	1.00	.01	.00; .03	25.56	17	.00	.00	+.01
3. Strong invariance	198.06	181	1.00	1.00	.02	.00; .03	44.08	33	.00	.00	+.01
4. Strict invariance	212.80	191	1.00	1.00	.02	.00; .03	15.45	10	.00	.00	.00
5. Latent variance-covariance invariance	206.14	194	1.00	1.00	.01	.00; .03	1.27	3	.00	.00	-.01
6. Latent mean invariance	244.40	197	.99	.99	.03	.01; .04	13.02*	3	-.01	-.01	+.02

Note. *$p < .01$; χ^2: Chi square test of model fit and associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker–Lewis Index; RMSEA: Root Mean Square Error of Approximation and 90% Confidence Interval (CI); Δ: Change according to the previous retained model; $\Delta\chi^2$: Chi square difference test calculated using the Mplus DIFFTEST option; [1] Additional tests of measurement invariance were conducted as a function of an age by ID level interaction. The results from these tests replicated those related to age and ID levels, without suggesting any interaction between age and ID level.

Table S4

Internal Consistency of Youth Self-Report per Characteristic.

	Prosocial behaviors (α)	Aggressive behaviors (α)
Sex		
Female	.77	.88
Male	.80	.83
Age (median split)		
Younger	.79	.84
Older	.80	.85
ID level		
Mild	.77	.86
Moderate	.80	.83
Country		
Canada	.80	.72
Australia	.78	.84

Table S5

Measurement Invariance Models for Teacher's Reports of Youth Social Behaviors.

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA
Measurement Invariance (Bifactor-CFA): Sex											
1. Configural invariance	278.69*	150	.98	.97	.08	.06; .09	–	–	–	–	–
2. Weak invariance	277.24*	177	.98	.98	.06	.05; .08	29.65	27	+.01	+.01	-0.02
3. Strong invariance	295.70*	203	.99	.98	.06	.04; .07	24.23	26	.00	.00	-0.01
4. Strict invariance	300.94*	218	.99	.99	.05	.04; .07	16.86	15	.00	.00	-0.01
5. Latent variance-covariance invariance	311.68*	221	.99	.99	.05	.04; .07	45.00*	3	.00	.00	.00
6. Latent mean invariance	319.52*	224	.98	.99	.06	.04; .07	6.79	3	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Age											
1. Configural invariance	253.57*	150	.98	.97	.08	.06; .09	–	–	–	–	–
2. Weak invariance	317.19*	177	.97	.97	.08	.07; .10	467.20*	27	-0.01	-0.01	+.01
3. Strong invariance	329.12*	203	.97	.97	.07	.06; .09	83.59*	26	+.00	+.01	-0.01
4. Strict invariance	373.94*	218	.97	.97	.08	.06; .09	56.76*	15	-0.01	.00	+.01
5. Latent variance-covariance invariance	311.32*	221	.98	.98	.06	.04; .07	4.08	3	+.01	+.01	-0.02
6. Latent mean invariance	336.25*	224	.98	.98	.07	.05; .08	12.45*	3	.00	.00	+.01
Measurement Invariance (Bifactor-CFA): ID Level											
1. Configural invariance	274.23*	150	.98	.97	.08	.07; .10	–	–	–	–	–
2. Weak invariance	301.98*	177	.98	.98	.07	.06; .09	49.42*	27	.00	.00	-0.01
3. Strong invariance	312.85*	203	.98	.98	.07	.05; .08	26.79	26	.00	+.01	-0.01
4. Strict invariance	356.85*	218	.98	.98	.07	.06; .08	41.31*	15	.00	.00	+.01
5. Latent variance-covariance invariance	348.55*	221	.98	.98	.07	.05; .08	11.52*	3	.00	.00	.00
6. Latent mean invariance	340.49*	224	.98	.98	.06	.05; .08	2.60	3	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Country											
1. Configural invariance	277.83*	150	.98	.97	.08	.06; .09	–	–	–	–	–
2. Weak invariance	329.41*	177	.98	.97	.08	.07; .09	65.55*	27	.00	.00	.00
3. Strong invariance	405.35*	203	.97	.97	.08	.07; .010	67.95*	26	-0.01	.00	+.01
4. Strict invariance	453.57*	218	.96	.97	.09	.08; .10	41.03*	15	-0.01	.00	.00
5. Latent variance-covariance invariance	400.73*	221	.97	.97	.08	.06; .09	12.91*	3	.00	+.01	-0.01
6. Latent mean invariance	428.24*	224	.97	.97	.08	.07; .09	15.52*	3	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Time											
1. Configural invariance	506.07*	351	.99	.98	.04	.03; .05	–	–	–	–	–
2. Weak invariance	514.35*	378	.99	.99	.03	.03; .04	33.46	27	.00	.00	.00
3. Strong invariance	536.62*	404	.99	.99	.03	.03; .04	27.00	26	.00	.00	.00
4. Strict invariance	591.98*	419	.98	.98	.04	.03; .04	41.57*	15	.00	.00	.00
5. Latent variance-covariance invariance	585.55*	422	.98	.98	.04	.03; .04	7.81	3	.00	.00	.00
6. Latent mean invariance	597.09*	425	.98	.98	.04	.03; .04	8.37	3	.00	.00	.00

Note. *$p < .01$; χ^2: Chi square test of model fit and associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker–Lewis Index; RMSEA: Root Mean Square Error of Approximation and 90% Confidence Interval (CI); Δ: Change according to the previous retained model; $\Delta\chi^2$: Chi square difference test calculated using the Mplus DIFFTEST option.

Table S6

Measurement Invariance Models for Parental Reports of Youth Social Behaviors

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA
Measurement Invariance (Bifactor-CFA): Sex											
1. Configural invariance	177.23	150	.99	.99	.05	.00; .07	–	–	–	–	–
2. Weak invariance	195.10	177	1.00	.99	.03	.00; .06	84.27*	27	.00	+.01	-.01
3. Strong invariance	226.92	203	.99	.99	.04	.00; .06	41.48	26	.00	.00	.00
4. Strict invariance	243.55	218	.99	.99	.04	.00; .06	46.63*	15	.00	.00	.00
5. Latent variance-covariance invariance	234.60	221	1.00	1.00	.03	.00; .05	50.52*	3	.00	.00	-.01
6. Latent mean invariance	257.26	224	.99	.99	.04	.00; .06	7.47	3	-.01	-.01	+.02
Measurement Invariance (Bifactor-CFA): Age											
1. Configural invariance	184.31*	150	.99	.98	.05	.02; .08	–	–	–	–	–
2. Weak invariance	229.73*	177	.98	.97	.06	.04; .08	48.81*	27	-.01	-.01	+.01
3. Strong invariance	254.03*	203	.98	.98	.06	.03; .08	24.01	26	.00	.00	.00
4. Strict invariance	278.94	218	.97	.98	.06	.04; .08	34.45*	15	.00	.00	.00
5. Latent variance-covariance invariance	246.45	221	.99	.99	.04	.00; .06	3.35	3	+.02	+.01	-.02
6. Latent mean invariance	255.90	224	.99	.99	.04	.00; .07	5.82	3	.00	.00	+.01
Measurement Invariance (Bifactor-CFA): ID Level											
1. Configural invariance	195.61*	150	.99	.98	.06	.03; .08	–	–	–	–	–
2. Weak invariance	235.16*	177	.98	.98	.06	.04 .08	54.21*	27	.00	.00	.00
3. Strong invariance	259.98*	203	.98	.98	.06	.03; .08	20.22	26	.00	.00	-.01
4. Strict invariance	301.37*	218	.97	.97	.07	.05; .08	43.58*	15	-.01	-.01	+.01
5. Latent variance-covariance invariance	247.77	221	.99	.99	.04	.00; .06	56.68*	3	+.02	+.02	-.03
6. Latent mean invariance	274.31	224	.98	.99	.05	.03; .07	10.58	3	-.01	-.01	+.01
Measurement Invariance (Bifactor-CFA): Country											
1. Configural invariance	170.19	150	.99	.99	.04	.00; .07	–	–	–	–	–
2. Weak invariance	216.40	177	.99	.99	.05	.02; .07	73.80*	27	-.01	-.01	+.01
3. Strong invariance	256.61*	203	.98	.98	.05	.03; .07	40.35	26	.00	.00	.00
4. Strict invariance	272.79*	218	.98	.98	.05	.03; .07	39.42*	15	.00	.00	.00
5. Latent variance-covariance invariance	281.75*	221	.98	.98	.06	.03; .07	.97	3	.00	.00	.00
6. Latent mean invariance	287.75*	224	.98	.98	.06	.04; .08	5.40	3	.00	.00	.00
Measurement Invariance (Bifactor-CFA): Time											
1. Configural invariance	430.07*	350	.98	.97	.03	.02; .04	–	–	–	–	–
2. Weak invariance	158.07	378	.98	.97	.03	.02; .04	31.54	27	.00	.00	.00
3. Strong invariance	480.24*	404	.98	.98	.03	.02; .04	21.66	26	.00	.00	.00
4. Strict invariance	492.43*	419	.98	.98	.03	.02; .04	16.44	15	.00	.00	.00
5. Latent variance-covariance invariance	490.41	422	.98	.98	.03	.01; .04	3.70	3	.00	.00	.00
6. Latent mean invariance	488.26	425	.98	.98	.03	.01; .04	1.45	3	.00	.00	.00

Note. *p < .01; χ^2: Chi square test of model fit and associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker–Lewis Index; RMSEA: Root Mean Square Error of Approximation and 90% Confidence Interval (CI); Δ: Change according to the previous retained model; $\Delta\chi^2$: Chi square difference test calculated using the Mplus DIFFTEST option.

Appendix A
Complete List of Items for the Measures of Youth Social Behaviors

Youth Self-Report	Teacher Report	Parental Report
English		
Over the last week…	Over the last month, this student…	Over the last month, my child has…
Prosocial Behavior		
P1 You have been attentive to other people's feelings (for example, when someone was not going well, you went to see him or her).	… was considerate of other people's feelings.	… been considerate of other people's feelings.
P2 You shared your belongings with others.	… shared readily with other students (for example food, games, pencils).	… readily shared something with other young people (for example books, games, food).
P3 You have been helpful when someone was hurt of feeling unwell.	… was helpful when someone got hurt, upset or feeling ill.	… been helpful if someone is hurt, upset or feeling ill.
P4 You have played with your friends.	… had at least one friend.	… had at least one good friend.
P5 *Not applicable*	… was generally liked by other students.	… been generally liked by other young people.
P6 *Not applicable*	… was kind to younger students.	… been kind to younger children.
P7 You helped others.	… often volunteered to help others (parents, teachers, other students).	… often volunteered to help others (parents, teachers, children).
Aggressive Behavior		
A1 You became physically aggressive when teased.	… reacted in an aggressive manner when teased (that is, slap, hit or push the other).	… reacted in an aggressive manner when teased (that is, slap, hit or push the other).
A2 You became physically aggressive when contradicted.	… reacted in an aggressive manner when contradicted (that is, slap, hit or push the other).	… reacted in an aggressive manner when contradicted (that is, slap, hit or push the other).
A3 You forced others to give you something you wanted by scaring them.	… scared other students to get what he/she wanted.	… scared other children to get what he/she wanted.
A4 You became physically aggressive or very angry because someone hurt you by accident.	…when someone accidentally hurt him/her (such as by bumping into him/her), he/she reacted with anger and violence.	…when someone accidentally hurt him/her (such as by bumping into him/her), he/she reacted with anger and violence.
A5 *Not applicable*	… when mad at someone, said bad things behind the other's back.	… when mad at someone, said bad things behind the other's back.
A6 *Not applicable*	… physically attacked other students.	… physically attacked people.
A7 You became physically aggressive when something was taken away from you.	… reacted in an aggressive manner when something was taken away from him/her (that is, slap, hit or push the other).	… reacted in an aggressive manner when something was taken away from him/her (that is, slap, hit or push the other).
A8 *Not applicable*	…hit, bitten or kicked other students.	…hit, bitten or kicked other children.
Never (0); 1 time (1); 2 times (2); 3 times (3); 4 times (4); 5 times or more (5); Doesn't understand the statement	Never (1); Rarely (2); Sometimes (3); Often (4); Very often (5).	Never (1); Rarely (2); Sometimes (3); Often (4); Very often (5).

Item	Youth Self-Report	Teacher Report	Parental Report
French			
	Au cours de la semaine dernière…	Au cours du dernier mois, cet élève…	Au cours du dernier mois, votre enfant…
Comportements prosociaux			
P1	Tu as été attentive(ve) aux autres, à ce qu'ils ressentent (par exemple, quand quelqu'un n'allait pas bien, tu es allé(e) le/la voir).	… a été attentif(ve) aux autres, tenu compte de ce qu'ils ressentent.	… a été attentif(ve) aux autres, tenu compte de ce qu'ils ressentent.
P2	Tu as partagé tes choses avec les autres.	… a partagé facilement avec les autres élèves (par exemple, friandises, jouets, crayons, etc.).	… a partagé facilement avec les autres enfants (par exemple, friandises, jouets, crayons, etc.).
P3	Tu as aidé quand quelqu'un s'est fait mal ou ne s'est pas senti bien.	… a volontiers apporté son aide quand quelqu'un s'est fait mal ou ne s'est pas senti bien.	… a volontiers apporté son aide quand quelqu'un s'est fait mal ou ne s'est pas senti bien.
P4	Tu as joué avec tes ami(e)s.	… a eu au moins un(e) ami(e).	… a eu au moins un(e) ami(e).
P5	*Not applicable*	… a généralement été aimé(e) par les autres élèves.	… a généralement été aimé(e) par les autres enfants.
P6	*Not applicable*	… a été gentil(le) avec les élèves plus jeunes.	… a été gentil(le) avec les enfants plus jeunes.
P7	Tu as aidé les autres.	… a toujours été prêt(e) à aider les autres [c'est-à-dire, parents, enseignant(e)s, autres élèves].	… a toujours été prêt(e) à aider les autres [c'est-à-dire, parents, enseignant(e)s, autres enfants].
Comportements agressifs			
A1	Tu es devenu(e) agressif(ve) physiquement lorsqu'un autre jeune t'as agacé(e).	… lorsqu'on le(la) taquinait, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).	… lorsqu'on le(la) taquinait, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).
A2	Tu es devenu(e) agressif(ve) physiquement parce que les autres n'étaient pas d'accord avec toi.	… lorsqu'on le(la) contredisait, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).	… lorsqu'on le(la) contredisait, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).
A3	Tu as fait peur aux autres pour les forcer à te donner quelque chose que tu voulais.	… a fait peur aux autres élèves afin d'obtenir ce qu'il(elle) voulait.	… a fait peur aux autres afin d'obtenir ce qu'il(elle) voulait.
A4	Tu es devenu(e) agressif(ve) physiquement ou très en colère parce qu'on t'a fait mal sans le faire exprès.	…lorsque quelqu'un lui faisait mal accidentellement [par exemple, en le(la) bousculant], il(elle) s'est fâché(e) et a commencé une bagarre (une chicane).	… lorsque quelqu'un lui faisait mal accidentellement [par exemple, en le(la) bousculant], il(elle) s'est fâché(e) et a commencé une bagarre (une chicane).
A5	*Not applicable*	… lorsqu'il(elle) était fâché(e) contre quelqu'un a dit de vilaines choses dans le dos de l'autre personne.	… lorsqu'il(elle) était fâché(e) contre quelqu'un, a dit de vilaines choses dans le dos de l'autre personne.
A6	*Not applicable*	… a attaqué physiquement les autres élèves.	… a attaqué physiquement les autres.
A7	Tu es devenu(e) agressif(ve) physiquement lorsqu'une personne t'a pris quelque chose.	… lorsqu'on lui prenait quelque chose, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).	… lorsqu'on lui prenait quelque chose, a réagi de façon agressive (par exemple, taper, pousser ou frapper l'autre).
A8	*Not applicable*	… a frappé, mordu, donné des coups de pied à d'autres enfants.	… a frappé, mordu, donné des coups de pied à d'autres enfants.
	Jamais (0); 1 fois (1); 2 fois (2); 3 fois (3); 4 fois (4); 5 fois ou plus (5); ne comprend pas l'énoncé	Jamais (1); Rarement (2); Parfois (3); Souvent (4); Très souvent (5)	Jamais (1); Rarement (2); Parfois (3); Souvent (4); Très souvent (5)