CrossMark

Reciprocal Relationships between Teacher Ratings of Internalizing and Externalizing Behaviors in Adolescents with Different Levels of Cognitive Abilities

Alexandre J. S. Morin [1] · A. Katrin Arens[2] · Christophe Maïano[3] · Joseph Ciarrochi[1] · Danielle Tracey[4] · Philip D. Parker[1] · Rhonda G. Craven[1]

Abstract Are internalizing and externalizing behavior problems interrelated via mutually reinforcing relationships (with each behavior leading to increases over time in levels of the other behavior) or mutually suppressing relationships (with each behavior leading to decreases over time in levels of the other behavior)? Past research on the directionality of these relationships has led to ambiguous results, particularly in adolescence. Furthermore, the extent to which prior results will generalize to adolescents with low levels of cognitive abilities remains unknown. This second limit is particularly important, given that these adolescents are known to present higher levels of externalizing and internalizing behaviors than their peers with average-to-high levels of cognitive abilities, and that the mechanisms involved in the reciprocal relationships between these two types of behaviors may differ across both populations. This study examines the directionality of the longitudinal relationships between externalizing and internalizing behavior problems as rated by teachers across three measurement waves (corresponding to Grades 8–10) in matched samples of 138 adolescents (34.78 % girls) with low levels of cognitive abilities and 556 adolescents (44.88 % girls) with average-to-high levels of cognitive abilities. The results showed that the measurement structure was fully equivalent across time periods and groups of adolescents, revealing high levels of developmental stability in both types of problems, and moderately high levels of cross-sectional associations. Levels of both internalizing and externalizing behaviors were higher among adolescents with low levels of cognitive abilities relative to those with average-to-high levels of cognitive abilities. Finally, the predictive analyses revealed negative reciprocal longitudinal relationships (i.e., mutually suppressing relationships) between externalizing and internalizing problems, a result that was replicated within samples of adolescents with low, and average-to-high levels of cognitive ability.

Keywords Internalizing behaviors · Externalizing behaviors · Adolescents · Cognitive abilities · Longitudinal · Autoregressive cross-lagged · Teacher ratings

Alexandre J. S. Morin and A. Katrin Arens contributed equally to this article and their order was determined at random: both should thus be considered first authors.

✉ Alexandre J. S. Morin
  Alexandre.Morin@acu.edu.au

[1] Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, NSW, Australia

[2] German Institute for International Educational Research, Frankfurt, Germany

[3] Université du Québec en Outaouais, Saint-Jérome, Québec, Canada

[4] Western Sydney University, Milperra, Australia

Introduction

Internalizing and externalizing behaviors represent two broad dimensions of behavioral problems, typically assessed in terms of severity, which respectively encompass depression, anxiety and social withdrawal (internalizing) or aggression, opposition and hyperactivity-inattention (externalizing) (e.g., Achenbach 2009; Krueger and Markon 2006). Despite forming distinct dimensions, the co-occurrence of internalizing and externalizing behaviors is well documented among youth (Gilliom and Shaw 2004; Krueger and Markon 2006; Reitz et al. 2005). However,

their temporal ordering remains unclear. Some research supports positive unidirectional relationships according to which externalizing behaviors increase later levels of internalizing behaviors (e.g., Curran and Bollen 2001; Van der Giessen et al. 2013), or internalizing behaviors predict later increases in externalizing behaviors (e.g., Herrenkohl et al. 2010; Vieno et al. 2008). In contrast, accumulating evidence suggests negative longitudinal relationships whereby each problem predicts decreases in the other problem, particularly in adolescence (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Lee and Bukowski 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016; Wiesner 2003). Using teacher ratings of behavioral problems, which are more likely to focus on observable behaviors than self-reports and less likely to suffer from self-consistency and memory biases, this study aims to clarify the temporal ordering of internalizing and externalizing behaviors. Furthermore, this study explores the extent to which the observed relationships generalize to samples of adolescents with average-to-high levels of cognitive abilities relative to those with low levels of cognitive abilities, for whom the mechanisms involved in the co-occurrence of internalizing and externalizing behaviors may differ.

Theoretical Perspectives on the Co-Occurrence of Internalizing and Externalizing Behaviors

The high level of co-occurrence between internalizing and externalizing behaviors in young populations has been well documented, with correlations generally close to, or above, .50 (e.g., Burt et al. 2008; Jackson et al. 2000). Many explain this co-occurrence by the presence of common determinants, and research supports the presence of shared determinants for internalizing and externalizing behaviors (e.g., Caron and Rutter 1991; Burt et al. 2008; Lee and Stone 2012; Moilanen et al. 2010). However, although shared determinants may explain the co-occurrence of internalizing and externalizing behaviors at a specific time point, it is unlikely to explain how each problem may be involved in the development of the other, particularly during adolescence given that the first onset of these behaviors typically occurs in childhood.

Mutually Reinforcing Relationships

Alternative theoretical explanations for the co-occurrence of internalizing and externalizing behaviors focus on the likely impact of one type of problem on the emergence of the other. First, Capaldi's (1991, 1992) *Failure Hypothesis* suggests that the presence of externalizing behaviors impinge on the development of competence in the social and academic areas, which precipitates failures experiences

in these areas. The accumulation of these failures experiences is proposed to lead to an increased risk of developing internalizing behaviors. Second, the *Acting Out Hypothesis* claims that youth may "mask" internalizing behaviors by acting out and engaging in externalizing behaviors (Carlson and Cantwell 1980; Gold et al. 1989), a phenomenon that may be amplified in childhood and adolescence when internalizing symptoms tend to be accompanied by increased levels of irritability (Wolff and Ollendick 2006). Even more direct mechanisms have been proposed through which internalizing behaviors increase the reactivity of neural regions associated with threat responses, leading to difficulties of emotion regulation in stressful situations and an increased likelihood of aggressive responses to these situations (Drabick et al. 2010).

Third, internalizing behaviors may also predict increased levels of externalizing behaviors through the action of mechanisms similar to those invoked by the *Failure Hypothesis*. Thus, internalizing behaviors are also associated with a more aversive interpersonal style, lower levels of social competence, and a tendency to withdraw from social interactions (Joiner and Coyne 1999), possibly leading to peer rejection and victimization (i.e., experiences of failures in the social area), two known predictors of externalizing behaviors (Marsh et al. 2011). Similarly, through their associations with difficulties of concentration, internalizing behaviors may also lead to failures in the academic area (Moilanen et al. 2010; Quiroga et al. 2013). As their experiences of failures in these two areas accumulate, youth may in turn be more likely to join deviant peer networks, further increasing their likelihood of developing externalizing behaviors (Connell and Dishion 2006; Oland and Shaw 2005).

Fourth, these unidirectional hypotheses have recently been integrated in Moilanen et al.' (2010) *Adjustment Erosion Hypothesis*. This hypothesis states that both internalizing and externalizing behaviors lead to lower levels of social and academic competence, resulting in accumulated failures, in turn resulting in increased risks for the other type of problem. Mutually reinforcing relationships are thus expected between internalizing and externalizing behaviors. It is noteworthy that the *Failure Hypothesis* and the *Adjustment Erosion Hypothesis* implicitly suggest that the reciprocal effects of each behavior on the other might increase with age due to the accumulation of failure experiences. In line with this possibility, a fifth model suggests that adolescence might represent a particularly critical period for investigating possible reciprocal relationships between internalizing and externalizing behaviors. Indeed, in a more comprehensive model hereafter referred to as the *Socio-Developmental Milestones Perspective*, Oland and Shaw (2005) explicitly propose that these reciprocal relationships should increase with time. More

precisely, this model proposes that externalizing behaviors should not lead to the development of internalizing behaviors as long as youth have not reached the socio-developmental milestone of developing cognitive abilities for self-evaluation, self-reflection, and empathy that are necessary to realize the negative impact of externalizing behaviors in their lives. This model also suggests that internalizing behaviors should not lead to the development of externalizing behaviors as long as youth do not develop strong intimate peer relationships. The absence of intimate relationships limits the possibility to experience negative interactions, negative feedback, and rejection, which may in turn reduce the risk of developing externalizing behaviors in reactions to these experiences. As the likelihood of attaining these socio-developmental milestones increases with age, so should the strength of the reciprocal relationships between internalizing and externalizing behaviors.

Mutually Suppressing Relationships?

So far, all theoretical models used in this line of enquiry have assumed positive predictive relationships between these two types of problems, and are thus unable to explain prior reports suggesting that these relationships may in fact be negative in adolescence (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Lee and Bukowski 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016; Wiesner 2003). Interestingly, alternative theoretical perspectives emerging from animal models, neuropsychological research, and from examination of the clinical characteristics of these behavior problems indirectly suggest that these relationships may also be negative, with one type of behavior predicting decreases in the other.

Thus, drawing parallels with the expanded fight, flight or freeze mechanisms involved in reactions to threatening situations (Bracha et al. 2004; Maack et al. 2015), internalizing behaviors are often assumed to correspond to the last two mechanisms, which are incompatible with the display of externalizing responses. Thus, anxiety is known to result in a priming of the typical fight-or-flight response in the anticipation of a threatening situation, but this priming typically results in decisions to avoid the anticipated source of stress (Kunimatsu and Marsee 2012; Zinbarg et al. 1992). As such, it is not surprising to note that avoidance is an important part of the clinical picture of anxiety (Maack et al. 2015; Zinbarg et al. 1992), and tends to be particularly marked in social situations (Miers et al. 2014). It is also worth noting that social withdrawal is considered to represent a key component of internalizing behaviors (Achenbach 2009; Rubin et al. 2009), and that recent research has also shown avoidance to be a key component of the clinical picture of depression (Ottenbreit et al. 2014). An even closer correspondence is assumed

between depression and the "freeze" response to threatening situations (Zinbarg et al. 1992). Thus, the *Social Competition Hypothesis* of depression (Price et al. 1994; Nesse 2000), which has been more recently expanded to include anxiety (Price et al. 2004), proposes that internalizing behaviors are an evolved version of primitive adaptive processes gone awry. Taking the example of ritualistic fighting for dominance, a behavior common in other mammalian species, this theory proposes that depression (and to a lesser extent anxiety) is similar to the automated yielding behaviors implicated in the "freeze" or "flight" response to stressful situations. As was the case for avoidance, these automated "yielding" responses are also incompatible with the display of externalizing responses.

Whereas the above considerations suggest that higher levels of internalizing disorders may possibly predict decreases in levels of externalizing behaviors, additional models and research results suggest that higher levels of externalizing disorders may also predict decreases in levels of internalizing behaviors. Thus, externalizing behaviors may represent a form of adaptation to troubled environments, or a way through which youth may seek alternate forms of success (Hayes and Ciarrochi 2015). Adopting an operant perspective on externalizing behaviors, Snyder and his colleagues (Snyder 2002; Snyder and Patterson 1995; Snyder and Stoolmiller 2002) noted that externalizing behaviors often serve an important function, being associated with a variety of positive and negative reinforcements. For instance, reacting with aggression to situations of coercion, conflict, or even bullying may help to end the aversive situation sooner. Similarly, externalizing behaviors are also known to lead to social interactions with deviant peers who tend to value and reinforce delinquent activities. These delinquent activities may provide short term experiences of "success", in terms of gaining attention from peers and adults. Altogether, these "desirable" outcomes of externalizing behaviors may compensate for their associated failure experiences, potentially leading to lower levels of internalizing behaviors. Still, it is important to keep in mind that the possible social benefits of externalizing behaviors may only appear during adolescence, based on recent research evidence suggesting that childhood externalizing behaviors rather tend to be associated with less positive social interactions with peers (Gooren et al. 2011; Van Lier and Koot 2010; Van Lier et al. 2012).

The Directionality of the Associations between Internalizing and Externalizing Behaviors

Testing the directionality of reciprocal relationships between two types of behavioral problems requires an autoregressive cross-lagged design, such as the one illustrated in Fig. 1, which includes three measurement points

Fig. 1 Autoregressive Cross-Lagged Model for Testing the Longitudinal Relationships between Internalizing and Externalizing Behaviors
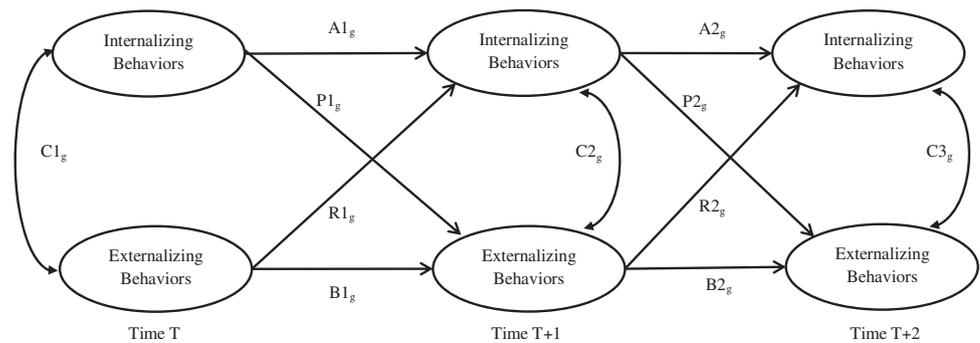
for consistency with the present study. As shown in Fig. 1, identifying the directionality of these associations requires the demonstration that internalizing behaviors at Time T predict externalizing behaviors at Time T + 1 (cross-lagged paths P1 and P2), and that externalizing behaviors at Time T predict internalizing behaviors at Time T + 1 (reciprocal cross-lagged paths R1 and R2), while controlling for the cross-sectional associations (paths C1–C3) and longitudinal stability (paths A1–A2 and B1–B2) (Mitchison et al. 2015). The numerical component (1, 2, 3) associated with each of these relationships (A, B, P, R, and C) indicates that they can be alternatively freely estimated or constrained to equality across time periods (as in tests of predictive equilibrium described later; Cole and Maxwell 2003). Conversely, the subscript $_g$ associated with each of these relationships indicates that they can be alternatively freely estimated or constrained to equality across groups of participants in the context of multi-group models.

Self-Reported Studies of Reciprocal Effects

So far, some studies have investigated bidirectional relationships between self-reported ratings of internalizing and externalizing behaviors using a scheme similar to Fig. 1, and relying mainly on samples of participants in their late childhood or adolescence. The results from these studies are inconsistent, with only a minority supporting direct or indirect positive bidirectional relationships, generally small in magnitude (Lee and Stone 2012) or limited to girls (Klostermann et al. 2015; Measelle et al. 2006). Other studies have supported direct or indirect unidirectional positive relationships going from internalizing behaviors to externalizing behaviors (Herrenkohl et al. 2010; Vieno et al. 2008), or from externalizing behaviors to internalizing behaviors, again limited to specific subpopulations (boys: Wiesner 2003; Klostermann et al. 2015; maltreated girls: Brentsilver et al. 2011). Finally, some studies failed to identify any cross-lagged relationships between internalizing and externalizing behaviors (Akse et al. 2007; Overbeek et al. 2001). In a comprehensive study involving four measurement points taken over two years in a sample of

adolescents, Wiesner (2003) reported a negative relationship whereby depression measured in wave 2 predicted decreases in delinquency 6 months later among girls. In a similar study of Korean fourth graders repeatedly measured over a 4 year period, Lee and Bukowski (2012) found that initial levels of externalizing behaviors predicted decreases over time in internalizing behaviors among boys and girls, whereas internalizing behaviors predicted decreases in externalizing behaviors among boys.

An important limitation of these studies is their exclusive reliance on self-reports (e.g., Van der Giessen et al. 2013), which leads to two potential limitations. First, self-reported data may lead to inflated estimates of stability (the autoregressive paths A1–A2 and B1–B2) due to self-consistency and memory biases whereby respondents recall their prior answers, or interpret their current behaviors in light of their past. As a result of potentially inflated stability estimates, it might become harder to predict change over time. Arguably, similar biases are likely to be present in any report of behavioral problems provided by the same informant over time. Teachers' report of youth behaviors as they occur in school over multiple school years may remedy this limit given that teachers change each year so that their assessments have to remain based on observed behaviors, rather than prior knowledge of past behaviors. A second limitation of self-report data is that it may lead to inflated estimates of the cross-lagged relationships between constructs. Although the focus of such assessments should be on behavioral problems of sufficient severity to be observable, youth knowledge of their own internal states may lead them to over-estimate some of the relationships. For instance, they are likely to be aware of the sadness or anxiety that accompany their delinquent activities, they may feel withdrawn from interactions with socially competent peers even if they externally appear well integrated into a delinquent peer group, or they may know that their acting out behaviors are attempts to cover up internalizing symptoms. In contrast, informant reports are more likely to focus on observable behaviors, allowing for a clearer understanding of which observable behavior leads to the other.

Informant-Reported Studies of Reciprocal Effects

We are aware of thirteen studies which have relied on informant reports. The first four focused on samples of children rated by their parents. Mathiesen et al. (2009) and Stone et al. (2015) failed to identify any form of cross-lagged relationships among parental ratings of young children's internalizing and externalizing behaviors, whereas Curran and Bollen (2001) found evidence of significant relationships between externalizing behaviors and later increases in internalizing behaviors among a sample of slightly older children. Boylan et al. (2010) relied on data from a sample of children initially aged 6–7 and followed biannually across three measurement waves. Their results revealed small positive relationships between maternal ratings of boys' oppositional symptoms and later increases in depressive symptoms, but this effect did not generalize across time waves. Among girls, the authors found evidence of negative relationships between mothers' ratings of depressive symptoms and later decreases in oppositional symptoms (again limited to a subset of time waves), supporting Wiesner's (2003) observation of negative relationships among girls. However, these authors also reported that the measurement model underlying mothers' ratings of depressive symptoms among girls was not invariant over time, casting doubts on the comparability of results across time waves.

The next four studies relied on a combination of self-reports and parental reports in samples of children and adolescents. Van der Giessen et al. (2013) assessed reciprocal relationships between self-reported ratings of depressive symptoms and parental ratings of adolescents' aggressive behaviors in a three wave longitudinal study with yearly intervals. The results revealed positive longitudinal relationships whereby aggressive behaviors predicted later increases in depressive symptoms, whereas the opposite relationship between depressive symptoms and later levels of aggressive behaviors was non-significant. Using data from a sample of at-risk boys measured at ages 6, 8, 10, 11, and 12 through parental reports and self-reports, Moilanen et al. (2010) reported similar relationships, albeit limited to a subset of time waves. Burt et al. (2008; for similar results obtained with the same sample, see Masten et al. 2005) tested the longitudinal relationships among ratings of internalizing and externalizing behaviors also obtained from a combination of parental reports and self-reports taken in early adolescence and 7, 10, and 20 years later. In this study, the only significant longitudinal relationship was a negative one between ratings of internalizing behaviors at age 17 and ratings of externalizing behaviors at age 20. Finally, Rogosch et al. (2010) followed a sample of 415 children initially aged between 7 and 9 years across four times points until they reached their late

adolescent years (15–18 years). In childhood, they relied on combined ratings of internalizing and externalizing behaviors provided by camp counsellor and teachers, whereas only self-reports were used in adolescence. This study revealed that internalizing behaviors in early childhood and early adolescence respectively predicted a decrease in externalizing behaviors in late childhood and late adolescence. These studies (Burt et al. 2008; Rogosch et al. 2010) support the presence of negative relationships between internalizing and externalizing behaviors (Boylan et al. 2010; Lee and Bukowski 2012; Wiesner 2003).

Apart from Rogosch et al. (2010), we are aware of five additional studies relying on teachers' ratings of children internalizing and externalizing behaviors, three of which relied on much younger samples than the present study. Van Lier and Koot (2010) annually followed a sample of children from kindergarten to Grade 4. Although their results revealed a direct positive relationship between teachers' ratings of externalizing behaviors in kindergarten and internalizing behaviors in Grade 1, none of the other direct relationships proved significant. In a second study of kindergartners followed over a two-year period, Gooren et al. (2011) also report evidence of a significant relationship between teachers' ratings of conduct disorders and later levels of depression, but show this relationship to be mediated by the effects of conduct disorders on peer rejection. No evidence of a reciprocal, direct or indirect, relationship between teacher's ratings of depressive symptoms and later levels of conduct disorders was evidenced. The third study (Van Lier et al. 2012) essentially replicated these findings among a sample of slightly older children aged between 6 and 8, which demonstrated an indirect relationship (with no evidence of direct relationship) between teachers' rating of externalizing behaviors at age 6 and internalizing behaviors at age 8 as mediated by lower level of academic achievement and higher levels of peer victimization at age 7. Again, no direct or indirect relationship was found between teachers' ratings of internalizing behaviors and later levels of externalizing behaviors.

Two other studies considered samples of children followed up into adolescence. Englund and Siebenbruner (2012) reported negative relationships between teachers' ratings of internalizing behaviors at age 7 and levels of externalizing behaviors at age 9. However, none of the reciprocal relationships between teachers' ratings of externalizing behaviors and later levels of internalizing behaviors, or of the later relationships between internalizing and externalizing behaviors (age 9–12 and 12–16) proved significant, which could possibly be explained by the small sample size used in this study (N = 191). Van der Ende et al. (2016) considered possible reciprocal relationships between internalizing and externalizing behaviors in adolescence, contrasting parents and teacher reports. Parental reports of

externalizing behaviors positively predicted later increases in parental reports of internalizing behaviors, whereas no evidence was found to support the reciprocal relationship between internalizing and externalizing behaviors when using parental reports. In contrast, teachers' reports also evidenced a negative relationship between ratings of internalizing behaviors and later levels of externalizing behaviors.

In sum, prior results obtained using the scheme presented in Fig. 1 remained inconsistent, suggesting that reciprocal relationships between internalizing and externalizing behaviors may be much smaller than previously assumed. Yet, accumulating evidence suggest that these relationships may be negative, particularly in research conducted among samples of adolescents or older children, and using informant reports (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Lee and Bukowski 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016; Wiesner 2003).

Predictive Equilibrium

When examining these prior results, it is noteworthy that among the few studies including more than two time waves, most studies in which significant relationships were observed found these relationships to be limited to a subset of the time intervals considered (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Klostermann et al. 2015; Masten et al. 2005; Moilanen et al. 2010; Rogosch et al. 2010; Van Lier and Koot 2010; Van Lier et al. 2012; Wiesner 2003). However, few studies directly and quantitatively tested whether the predictive relationships could be considered to be different, or equivalent, across time waves—what is commonly referred to as a test of predictive equilibrium (Cole and Maxwell 2003; Mitchison et al. 2015). The two studies that tested the predictive equilibrium of the relationships supported their generalizability across three time waves (Van der Ende et al. 2016; Van der Giessen et al. 2013), suggesting that such tests might be important. Thus, it remains unknown whether the observed discrepancies across time waves are related to random sampling variations around a common set of population parameters, or whether these discrepancies reflect meaningful developmental differences. Statistically, a predictive model that has reached equilibrium has the advantage of being more parsimonious (i.e., resulting in the free estimation of fewer time-invariant predictive paths), thus maximising the statistical power of the analyses and the stability of the estimation process.

More precisely, when examining the reciprocal relationships between internalizing and externalizing behaviors, there are three distinct elements to consider in the assessment of the predictive equilibrium of the system (Cole and

Maxwell 2003; Mitchison et al. 2015). A first test concerns whether the stability of internalizing and externalizing problems vary across time periods (i.e., whether the autoregressive paths A1 and B1 estimated between T and T + 1 are equal to the same paths A2 and B2 estimated across T + 1 to T + 2). So far, research shows that both types of problems tend to present high levels of stability and chronicity (Beyers and Loeber 2003; Burt et al. 2008; Van der Ende et al. 2016; Van der Giessen et al. 2013). There is, however, also evidence of time variation in the stability of internalizing and externalizing behaviors, with studies suggesting that stability may increase with age (Buist et al. 2004; Reitz et al. 2005).

A second test concerns whether cross-lagged relationships between internalizing and externalizing behaviors vary across time (i.e., whether cross-lagged paths P1 and R1 estimated between T and T + 1 are equal to the same paths P2 and R2 estimated across T + 1 to T + 2). This test verifies whether the impact of one problem on the other changes as a function of time. For example, the *Failure Hypothesis* and the *Adjustment Erosion Hypothesis* both suggest that the reciprocal effects of each types of behavior on the other might increase with age due to the accumulation of failure experiences, a possibility that appears supported by Wiesner (2003). Similarly, Oland and Shaw's (2005) *Socio-Developmental Milestones Perspective* explicitly suggest that the reciprocal relationships between internalizing and externalizing behaviors should become more marked during adolescence.

A third test concerns whether the time-specific correlations between internalizing and externalizing behaviors differ across time periods (i.e., whether the correlation C1 estimated at Time T is equal to the correlation C2 and C3 estimated at Time T + 1 and T + 2). This test verifies whether co-occurrence due to the impact of shared determinants increases or decreases over time. So far, research suggests that co-occurrence remains relatively stable across developmental periods (Beyers and Loeber 2003; Reitz et al. 2005), consistent with the idea that it reflects the action of shared determinants.

Generalizability to Youth with Low Levels of Cognitive Abilities

Research has demonstrated that youth with low levels of cognitive abilities present a higher likelihood of exhibiting internalizing and externalizing behaviors than their peers with average-to-high levels of cognitive abilities, making it particularly important to study the mechanisms involved in the development of these behaviors in this population (Dekker et al. 2002; Metcalfe et al. 2013; Weeks et al. 2014; Whitaker and Read 2006). These higher levels of internalizing and externalizing behaviors are not surprising

considering that risk factors for these behaviors also tend to be more prevalent among youth with low levels of cognitive abilities than among their peers with average-to-high levels of cognitive abilities, including more limited social skills, as well as higher levels of exposure to stressful life conditions, peer rejection, and victimization (Rose et al. 2011; Valas 1999; Wehmeyer 2005). However, although numerous studies have demonstrated higher levels of internalizing and externalizing behaviors among youth with low levels of cognitive abilities, no research has yet examined the reciprocal relationships between these behaviors in this population. This limitation is important as it restricts the extent to which knowledge gained in the far more numerous studies conducted among youth with average-to-high levels of cognitive abilities can be generalized to those with low levels of cognitive abilities.

As reviewed above, the broader research literature aiming to explain the relationships between internalizing and externalizing behaviors typically relies on the *Failure Hypothesis* (Capaldi 1991, 1992) and the *Adjustment Erosion Hypothesis* (Moilanen et al. 2010). These models rely on the assumption that each type of behavioral problem is likely to increase the risk of experiencing failures in a variety of domains, which in turn should lead to increased risks of also developing the other type of problem. Of particular relevance to youth with low levels of cognitive abilities, the *Dual-Risk Model* (Sameroff 1983) proposes that vulnerability traits, such as low levels of cognitive abilities, tend to increase the risk of experiencing problematic behavioral outcomes when exposed to environmental stressors. With the higher prevalence of internalizing and externalizing behaviors among youth with low levels of cognitive abilities, and their more frequent experiences of stresses (or "*failures*") in both the social and educational areas (e.g., Craven et al. 2015), the combination of these hypotheses suggest stronger reciprocal relationships between internalizing and externalizing behaviors among youth with low levels of cognitive abilities than among their peers with average-to-high levels of cognitive abilities.

In contrast, the more limited cognitive and social skills of youth with low levels of cognitive abilities are also likely to interfere with the mechanisms proposed to be at play in explaining possible reciprocal relationships between both types of behavioral problems. Thus, Oland and Shaw (2005) suggest that reciprocal relationships between internalizing and externalizing behaviors are conditional on the attainment of specific socio-developmental milestones. The cognitive deficits that characterize youth with low levels of cognitive abilities makes it less likely for them to successfully attain these developmental milestones, and may even lead them to develop altogether different definitions of "success" (e.g., Weeks et al. 2014). These observations suggest weaker reciprocal relationships between internalizing and externalizing behaviors among youth with low levels of cognitive abilities than among youth with average-to-high levels of cognitive abilities. Similarly, even though internalizing behaviors are known to lead to higher levels of social rejection, withdrawal, and victimization (Joiner and Coyne 1999), such experiences are already far more frequent for youth with low levels of cognitive abilities (Rose et al. 2011; Valas 1999; Wehmeyer 2005), and thus less likely to be further increased by internalizing behaviors. For youth with low levels of cognitive abilities already experiencing heightened social and educational difficulties, the power of these experiences to further enhance internalizing or externalizing behaviors (as proposed in the *Failure Hypothesis*, *Adjustment Erosion Hypothesis*, and *Dual-Risk Model*) may thus be minimal, suggesting weaker relationships than among youth with average-to-high levels of cognitive abilities.

Given that the co-occurrence between internalizing and externalizing behaviors is associated with greater impairment in terms of psychological, physiological, and social wellbeing (Fanti and Henrich 2010; Newman et al. 1998), which already tends to be compromised among youth with low levels of cognitive abilities (Helps 2015; Kiddle and Dagnan 2011), understanding the interplay between both types of problems appears to be of critical importance in this population. In response to the dearth of robust empirical investigation to evaluate the mechanisms at work for this vulnerable population, such investigations become even more paramount.

The Present Study

The first aim of the present study is to provide further insights on the nature of the longitudinal associations between teachers' ratings of adolescents' levels of observable internalizing and externalizing behaviors as they occur over three consecutive years. The theoretical perspectives typically invoked to explain how internalizing and externalizing behaviors may be reciprocally related over time (i.e., the *Failure Hypothesis*, the *Acting Out Hypothesis*, the *Adjustment Erosion Hypothesis,* and the *Socio-Developmental Milestones Perspective*) assume mutually reinforcing relationships between these types of behaviors whereby each one predicts increases in the other over time. A number of studies have provided tentative support to these hypotheses, showing either bidirectional (e.g., Lee and Stone 2012), or unidirectional (e.g., Curran and Bollen 2001; Herrenkohl et al. 2010; Van der Giessen et al. 2013) positive relationships among internalizing and externalizing behaviors, although additional studies failed to find evidence of significant relationships (e.g., Akse et al. 2007; Mathiesen et al. 2009; Overbeek et al. 2001). However,

accumulating evidence, particularly from studies relying on informant reports of internalizing and externalizing behaviors conducted among samples of adolescents, suggest that these relationships might be negative (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Lee and Bukowski 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016; Wiesner 2003). Although theoretical perspectives emerging from biological (fight-flight-freeze) or clinical (avoidance) research indirectly explain these unexpected negative relationships, these alternative models have never been used to guide research in this area. Furthermore, a key limitation of prior research is a frequent lack of consideration of the extent to which the relationships generalize to the various time intervals considered through systematic tests of predictive equilibrium. This limitation could have restricted the ability of these prior studies to detect significant relationships, and makes it impossible to determine whether any observed discrepancy reflects meaningful developmental differences rather than random sampling variations. Thus, although both theory and research allows us to expect significant reciprocal longitudinal relationships among teachers' ratings of adolescents' levels of internalizing and externalizing behaviors, the direction of these relationships remains an open research question due to the accumulation of contradictory research evidence. In sum, we consider the following two research questions:

Will the reciprocal relationships between teachers' ratings of adolescents' levels of internalizing and externalizing behaviors be mutually reinforcing (i.e., positive) or mutually suppressing (i.e., negative) (Research Question 1)?

To what extent will the reciprocal relationships between teachers' ratings of adolescents' levels of internalizing and externalizing behaviors generalize across the two time intervals considered between the three annual waves of measurement of the present study (Research Question 2)?

The second aim of this study is to assess whether the longitudinal relationships between teachers' ratings of adolescents' levels of internalizing and externalizing behaviors generalize across matched samples of adolescents with low levels of cognitive abilities and average-to-high levels of cognitive abilities. Although prior research leads us to expect higher levels of internalizing and externalizing behaviors among adolescents with low levels of cognitive abilities compared to their peers with average-to-high levels of cognitive abilities, evidence regarding the associations between these two types of behavioral problems is lacking among populations of youth with low levels of cognitive abilities. From a theoretical perspective, differential relationships can be expected among these two groups of adolescents. However, the nature of these differences remains again an open research question. Indeed, whereas some theoretical frameworks (i.e., the *Failure Hypothesis*, the *Adjustment Erosion Hypothesis*, and the *Dual-Risk Model*) suggest stronger reciprocal relationships between internalizing and externalizing behaviors among adolescents with low levels of cognitive abilities when compared to their peers with average-to-high levels of cognitive abilities, other perspectives suggest weaker reciprocal relationships (e.g., the *Socio-Developmental Milestones Perspective*). In sum, we consider the following research question and hypothesis:

Adolescents with low levels of cognitive abilities will present higher mean levels of internalizing and externalizing behaviors than those with average-to-high levels of cognitive abilities (Hypothesis 1).

To what extent will the reciprocal relationships between teachers' ratings of adolescents' levels of internalizing and externalizing behaviors generalize across matched samples of adolescents with low versus average-to-high levels of cognitive abilities (Research Question 3)?

Method

Sample, Procedure, and Matching

This study relies on a sample drawn from the Wollongong Youth Study, which was conducted in a number of regular secondary schools from the same Catholic Diocese and located in the regional and metropolitan areas of Wollongong (New South Wales, Australia). In 2003, all adolescents attending Grade 7 in the participating schools were targeted for participation and followed annually thereafter, resulting in a total sample of $N = 979$ adolescents (503 males; 476 females; 12–14 years old; M_{age} 12.41). We focus on three measurement waves (Grades 8, 9, and 10) for which teachers' ratings of internalizing and externalizing behaviors were available. Socioeconomic indicators, such as family occupation, structure, and first language closely match National Australian trends at the time of the study as reported by the Australian Bureau of Statistics (2005), supporting the idea that this sample is representative of the Australian population (Ciarrochi et al. 2012). This longitudinal study received annual approval from the university research ethics committee. Parents and adolescents also provided informed consent on an annual basis. For additional details on the Wollongong Youth Study, see Ciarrochi et al. (2012) and Heaven et al. (2009).

Upon entering secondary schools (in Grade 7), all students were required by the Department of Education and Training to complete two standardized measures of verbal and numerical aptitudes. The first test called "English Language and Literacy Assessment (ELLA)" assesses verbal aptitudes in writing, reading, and language ($\alpha = .87$),

while the second test called "Secondary Numeracy Assessment Program (SNAP)" assesses numeracy aptitudes in number, measurement, space, data, and numeracy problem-solving ($\alpha = .95$). Although these tests are not specifically designed to assess intelligence (IQ), similar tests of cognitive abilities are known to underpin a common global (g) factor (Deary et al. 2007; Frey and Detterman 2004). Furthermore, these specific tests have been found to be significantly related to the abbreviated Wechsler Scale of Intelligence (Heaven et al. 2011), to significantly predict future academic performance (Heaven, and Ciarrochi 2008), and to represent valid proxy measures of global IQ (Ciarrochi et al. 2012; Heaven and Ciarrochi 2012; Heaven et al. 2011; also see Weeks et al. 2014). In order to select the subsample of participants with low levels of cognitive abilities, we retained all adolescents who scored in the lowest 15 % on the score distribution of the ELLA and SNAP tests. The 15th percentile was selected as the cut-off point as (a) prevalence estimates of students with low levels of cognitive abilities in mainstream settings in Australia fall between 12–16 % (OECD 1999), and (b) 15 % of the population have an IQ that falls one standard deviation below the mean IQ score (Wechsler et al. 2003).

This procedure resulted in a sample of $N = 138$ adolescents with low levels of cognitive abilities (90 boys; 48 girls; 12–14 years old; $M_{age} = 12.41$). Most (79.7 %) were from an Anglo-Australian background. In terms of socio-economic status, 7.4 % had parents corresponding to an underclass level (e.g., retired, unemployed), 39.0 % to a working class level, 44.0 % to an intermediate social class level, and 9.6 % to a salariat (i.e., white collar professional, managers) social class level. These adolescents are best described as performing below or at national minimum standards in literacy and numeracy (Australian Curriculum, Assessment and Reporting Authority 2015). The higher proportion of males is not surprising as it is well documented that males are more likely to be identified as presenting learning difficulties than females (Australian Bureau of Statistics 2012; Westwood and Graham 2000).

Using the matchit package implemented in R (Ho et al. 2011; R Core Team 2013), we extracted a matched comparison sample of $N = 556$ adolescents with average-to-high levels of cognitive abilities (312 boys; 244 girls; 12–14 years old; $M_{age} = 12.41$; 85.6 % native English speakers) from the rest of the Wollongong Youth Study sample. For purposes of the present study, exact matching was conducted on the basis of school, gender, age in Grade 7, and first language. To retain as much information as possible, we matched up to six participants with average-to-high levels of cognitive abilities to a single comparable participant with low levels of cognitive abilities. This 6:1 matching resulted in a set of weights that were used in all models so that the sample reflected the matching procedure

(see Thoemmes and Kim 2011, for a review of propensity score matching procedures). These weights are such that their sum in each group is equal to the adolescents with average-to-high levels of cognitive abilities. Most (82.0 %) of these participants were from an Anglo-Australian background. In terms of socio-economic status, 4.2 % had parents corresponding to an underclass level, 28.6 % to a working class level, 49.2 % to an intermediate level, and 18.0 % to a salariat level. Because age, gender, and first language were used in the matching procedure, it is not surprising to note that these two samples did not differ from one another in terms of age, gender, and ethnic minority status ($p \geq .05$). However, the sample of adolescents with average-to-high levels of cognitive abilities tended to come from a slightly higher socio-economic background ($M = 1.97$) than their peers with low levels of cognitive abilities ($M = 1.79$; $p \leq .05$).

Instruments

Demographics

At the beginning of the study, participants self-reported their age (in unit increments) and gender (coded 0 for males and 1 for females). They also reported the main occupation of both of their parents, as well as the language they spoke most often at home, their birth country, and the birth country of their parents. All participants who reported having been born in Australia, speaking English at home, and having at least one parent also born in Australia were coded as being from an Anglo-Australian background (0) whereas the remaining participants were coded as being from an ethnic minority background (1). Finally, parental occupations were classified as a function of the major groupings of the Australian and New Zealand Standard Classification of Occupations (Australian Bureau of Statistics 2006). These were then simplified into broad categories of salariat (coded 3: including white collar professionals, managers, etc.), intermediate (coded 2: including trades, advanced clerical, service, etc.), working (coded 1: including elementary clerical or service workers, laborers, production, etc.), and underclass (coded 0: including unemployed, retired, homemakers, and pensioners; Jackson et al. 2012). These ratings were averaged across parents into a single measure of socio-economic status.

Internalizing and Externalizing Behaviors

At the end of each school year, homeroom teachers provided ratings of adolescents' internalizing and externalizing behaviors. Homeroom teachers meet with the target adolescent on a daily basis and are in charge of monitoring their school life across disciplines in a specific school year.

Homeroom teachers change across school years for any given adolescent. The items used for these ratings originate from the teacher version of the Multidimensional Peer Nomination Inventory (Pulkkinen et al. 1999). *Internalizing behaviors* were assessed by seven items ($\alpha = .83$–$.84$ across time waves) covering symptoms of depression (e.g., "Is sad and depressed") and anxiety (e.g., "Is shy in front of other students"). *Externalizing behaviors* included 13 items ($\alpha = .91$–$.93$ across time waves) covering direct (e.g., "Teases smaller and weaker students") and indirect (e.g., "Spreads rumors about other people's personal matters when he/she is mad at them") aggression, and symptoms of hyperactivity (e.g., Talks all the time) and inattention (e.g., "Is unable to concentrate on anything"). Ratings were given on a 4-point scale ranging from 0 (indicating that a specific behavior has not been observed) to 3 (indicating that the behavior describes the student very well). In this study, global scales of internalizing and externalizing behaviors are used.

A number of studies support the reliability and validity of these measures. Pulkkinen et al. (1999) showed that teachers' ratings of boys' and girls' (aged 12) internalizing and externalizing behaviors on this measure presented a satisfactory level of scale score reliability ($\alpha = .93$–$.94$ for externalizing and $.77$–$.82$ for internalizing), factor validity, and concurrent validity with parent and peer ratings of the same variables. Pulkkinen et al. (1999) also showed that these teachers' ratings were related in theoretically expected ways with gender (e.g., with boys scoring higher on externalizing behaviors), and were stronger predictors of peer reports of internalizing and externalizing behaviors than parent ratings. More recently, Heaven et al. (2009) demonstrated that teacher's ratings on these measures presented satisfactory levels of scale score reliability ($\alpha = .88$–$.93$), were moderately stable over time (Grades 7–11), and presented concurrent validity with adolescents self-reported ratings of antisocial personality. Ciarrochi et al (2007) also provided evidence of scale score reliability ($\alpha = .83$–$.93$), as well as of concurrent and predictive validity for teachers' ratings on this instrument in relationship with adolescents' (aged 12–14) self-reported ratings of hope, self-esteem, and positive attributional style.

Analyses

All models were estimated, while incorporating sample weights, using the robust weight least-square estimator (WLSMV) available in Mplus 7.3 (Muthén and Muthén 2014) to account for the ordered categorical nature of the four-point responses scales used for teacher ratings (Finney and DiStefano 2013; Rhemtulla et al. 2012). WLSMV estimation thus make no assumption regarding the underlying normality of the response scales, and results in the estimation of normally distributed latent factors, which represent the key variables of interest in the present study.

In this study, teachers competed an average 2.67 annual ratings of the participants, with 77.4 % of the participants having access to 3 sets of annual ratings, 12.1 % of the participants having access to 2 sets of annual ratings, and 10.5 % of the participants having access to 1 set of annual ratings, for a total of 1852 sets of time specific observations. A total of 637 participants had access to teacher's ratings in Grade 8, 613 in Grade 9, and 577 in Grade 10. Comparable figures for participants with average-to-high levels of cognitive abilities are: $M = 2.66$; 3 sets: 77.0 %; 2 sets: 12.2 %; 1 set: 10.8 %; total: 1480; Grade 8: 505; Grade 9: 491; Grade 10: 462. Comparable figures for participants with low levels of cognitive abilities are: $M = 2.70$; 3 sets: 79.0 %; 2 sets: 11.6 %; 1 set: 9.4 %; total: 372; Grade 8: 132; Grade 9: 122; Grade 10: 115. Once missing time waves are taken into account, none of the participants had more than one time-specific missing response at the item level. To account for missing responses, models were estimated based on the full available information, based on algorithms implemented in Mplus for WLSMV (Asparouhov and Muthén 2010). This procedure has comparable efficacy to multiple imputation, while being more efficient (Enders 2010; Larsen 2011), and allows missing data to be conditional on all variables included in the model.

The main predictive model estimated in this study is illustrated in Fig. 1, where the ovals represent the latent constructs estimated directly from their items. This model will be estimated as a multi-group model across samples of participants with low, versus average-to-high, levels of cognitive abilities. The results from this model will be used to address Research Question 1. The measurement part from this model was specified as invariant across groups and time waves based on the results from preliminary CFA models reported in the Appendix, which helps to maximize the parsimony, stability, statistical power, and stability the model. Latent means estimated as part of these preliminary CFA models will further be used to test Hypothesis 1. In the estimation of multiple-group longitudinal models, it is important to assess whether the predictive system has reached equilibrium (Cole and Maxwell 2003; Mitchison et al. 2015). These tests verify whether the overall pattern of associations between internalizing and externalizing behaviors generalize across time periods. We started with the estimation of a model in which all predictive paths and correlations were freely estimated across time and groups. From this model, invariance constraints were progressively added to the autoregressive paths (A and B), the cross-lagged paths (P and R), and the time-specific correlations (C). For all tests of invariance, we started with a conservative model in which constraints were imposed across groups and time points, before moving to less demanding

models of invariance across groups or time when results failed to support complete invariance assumptions. The results from these tests will further be used to address Research Questions 2 (equivalence of the predictions across time intervals) and 3 (equivalence of the predictions across groups). These models were then re-estimated while controlling for the effects of age, gender, minority status, and socio-economic status on the estimated relationships.

Model fit was evaluated using typical indicators as implemented in conjunction with WSLMV estimation (Hu and Bentler 1999; Yu 2002): (a) the chi-square statistic (χ^2), (b) the comparative fit index (CFI), (c) the Tucker-Lewis index (TLI), and (d) the root mean square error of approximation (RMSEA) and its 90 % confidence interval. According to typical interpretation guidelines (Hu and Bentler 1999; Marsh et al. 2005; Marsh et al. 2004; Vandenberg and Lance 2000) values greater than .90 and. 95 for the CFI and TLI respectively reflect adequate and excellent fit, while values smaller than .08 or .06 for the RMSEA respectively indicate acceptable and excellent fit. For the comparison of nested models used in the tests of measurement invariance (see Appendix) and predictive equilibrium, we examined changes in these fit indices, based on the recommended guidelines (Chen 2007; Cheung and Rensvold 2002; Vandenberg and Lance 2000) that a more constrained (or invariant) model can be considered as providing an equivalent fit than a less constrained model when it is accompanied by CFI or TLI decline of .010 or less and by a RMSEA increase of .015 or less. However, given that these recommended guidelines have been so far only validated for tests of measurement invariance, their relevance to tests of predictive invariance remains unknown. Thus, for tests of predictive invariance, we consider any decrease in model fit to suggest possible non-equivalence of the predictive paths, and combine examination of changes in goodness-of-fit indices with a more attentive consideration of parameter estimates. It should be noted that with WLSMV, the χ^2 values are "*estimated*" as the closest integer necessary to obtain a correct *p*-value, making it possible for the χ^2 and CFI values to be non-monotonic with model complexity. These apparent increases in fit should simply be interpreted as reflecting equally-fitting models.

Results

The results from the preliminary measurement models reported in the Appendix supported the adequacy of the a priori longitudinal measurement models, their strict invariance over time and groups of participants with low and average-to-high levels of cognitive abilities (i.e., the comparability of the ratings over time and groups), and the composite reliability of the latent constructs. Latent correlations also showed that both internalizing ($r = .466–.505$) and externalizing ($r = .452–.639$) behaviors were quite stable over time. In contrast, longitudinal correlations between internalizing and externalizing behaviors remained small, and often non-significant, although they tended to increase at later time waves. These results further supported Hypothesis 1 in showing that, when latent means were fixed to zero and latent variances to one for identification purposes for Grade 8 adolescents with average-to-high levels of cognitive abilities, adolescents with low levels of cognitive abilities presented higher levels of internalizing (+.520, $p \leq .01$) and externalizing (+.322, $p \leq .01$) behaviors in Grade 8 (expressed in standard deviation units). The magnitude of these differences remained stable over time (internalizing: +.450 and +.458 SD units in Grades 9 and 10; externalizing: + .261–.277 SD units in Grades 9 and 10). Finally, for both groups of adolescents, average levels of externalizing behaviors remained stable over time (non-significantly different from the grand mean of 0), while average levels of internalizing behaviors slightly decreased (−.259 SD per year, $p \leq .01$).

Starting from the final model of strict measurement invariance across groups and time points to ensure the complete comparability of teacher's ratings of internalizing and externalizing behaviors, we first estimated an auto-regressive cross-lagged predictive model (corresponding to Fig. 1) in which all predictive paths and time-specific correlations were freely estimated across groups and time (Model 1). The goodness-of-fit of all predictive models is reported in Table 1. These results show that Model 1 provided an acceptable level of fit to the data according to CFI and TLI ≥ .900 and an excellent fit to the data according to a RMSEA ≤ .060. From this model, the autoregressive (Model 2) and cross-lagged paths (Model 3) were constrained to be equal across groups and time waves. For both models, fit indices including a correction for parsimony (TLI, RMSEA) improved as a result of these constraints (ΔTLI = +.003–+.007; ΔRMSEA = −.001), while the CFI showed no decrease. These results thus support that the predictive system has reached equilibrium across time waves, and is fully equivalent across groups. However, when time-specific correlations were also constrained to be equivalent across groups and time points (Model 4), all goodness-of-fit indices revealed a slight decrease (ΔCFI = −.002; ΔTLI = −.002) accompanied by a significant chi-square difference test. Although this decrease remained relatively small, it calls into question the equivalence of the time-specific correlations across groups and time points. An examination of these time-specific correlations obtained in the previous model (Model 3) suggests that the time-specific correlations may not be fully equivalent across time waves. To verify this possibility, two models were estimated,

Table 1 Goodness-of-Fit Indices for the Predictive Models

	χ^2	df	CFI	TLI	RMSEA	RMSEA 90 % CI	CM	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA
Predictive models (no controls)												
1 Free Estimation across Time and Groups	5145.177*	3608	.914	.916	.036	[.033; .038]	–	–	–	–	–	–
2 Invariant Autoregressions across Time and Groups	5095.555*	3614	.917	.919	.035	[.033; .037]	1	7.819	6	+.003	.003	−.001
3 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups	4982.297*	3620	.924	.926	.034	[.031; .036]	2	4.132	6	+.007	+.007	−.001
4 Invariant Autoregressions, Cross-Lagged Paths, and Correlations across Time and Groups	5018.706*	3625	.922	.924	.034	[.032; .036]	3	27.431*	5	−.002	−.002	.000
5 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups; Inv. Correlations across Groups only	4933.753*	3623	.927	.929	.033	[.031; .035]	3	7.041	3	+.003	+.003	−.001
6 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups; Inv. Correlations across Time only	5021.047*	3624	.922	.924	.034	[.032; .036]	3	23.781*	4	−.002	−.002	.000
Predictive models (with controls)												
7 Free Estimation across Time and Groups	5699.599*	4072	.910	.911	.034	[.032; .036]	–	–	–	–	–	–
8 Invariant Effects of Controls across Time and Groups	5705.478*	4080	.910	.911	.034	[.032; .036]	7	15.297	8	.000	.000	.000
9 Invariant Autoregressions across Time and Groups	5648.935*	4086	.914	.915	.033	[.032; .036]	8	7.718	6	+.004	+.004	−.001
10 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups	5541.473*	4092	.920	.921	.032	[.030; .034]	9	4.262	6	+.006	+.006	−.001
11 Invariant Autoregressions, Cross-Lagged Paths, and Correlations across Time and Groups	5615.043*	4097	.916	.917	.033	[.031; .035]	10	41.951*	5	−.004	−.004	+.001
12 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups; Inv. Correlations across Groups only	5540.554*	4095	.920	.921	.032	[.030; .034]	10	7.688	3	.000	.000	.000
13 Invariant Autoregressions and Cross-Lagged Paths across Time and Groups; Inv. Correlations across Time only	5600.035*	4096	.917	.918	.033	[.030; .035]	10	34.058*	4	−.003	−.003	+.001

χ^2 WLSMV chi square test of exact fit, *df* degrees of freedom, *CFI* comparative fit index, *TLI* Tucker-Lewis index, *RMSEA* root mean square error of approximation, *CI* confidence interval, *CM* nested comparison model (typically the previous model in the sequence), $\Delta\chi^2$ chi-square difference test, Δ change relative to the comparison model (CM)

* $p \leq .01$

one in which the time-specific correlations were constrained to be equivalent across groups (Model 5) but not time waves, and one in which they were constrained to be equivalent across time waves, but not groups (Model 6). The goodness-of-fit indices associated with Model 6 mimic those obtained for Model 4, suggesting that the time-specific correlations differ across time waves. In contrast, Model 5 supports the equivalence of the time-specific correlations across groups. Model 5 was thus retained as the final model. The results from the models including the demographic controls (Models 7–13) lead to identical conclusions, supporting the superiority of Model 12 in which all regressive paths are constrained to equality across groups and time wave, whereas the time-specific correlations are allowed to differ across time waves, but not across groups.

To answer Research Questions 2 and 3, these results show that the predictive relationships across (paths P and R in Fig. 1) and within (the paths A and B in Fig. 1) constructs were fully equivalent across the two sets of time intervals considered in this study (Grades 7–8 and Grades 8–9), as well as across groups of participants with average-to-high, and low, levels of cognitive abilities. However, they also show that although the time-specific correlations among constructs (C in Fig. 1) were equivalent across groups of participants with average-to-high, and low, levels of cognitive abilities, they tend to change over time.

Parameter estimates for Models 5 and 12 are reported in Table 2. The results from models excluding (Model 5) or including the demographic controls (Model 12) are essentially identical, supporting the idea that these controls were not necessary. We thus focus on the results from Model 5 (excluding controls). Standardized estimates from this model are also reported in Fig. 2. These results first show moderately high (and similar), levels of longitudinal stability in levels of internalizing ($\beta = .566$ to $.654$) and externalizing ($\beta = .598$ to $.692$) behaviors in both groups. The size of these coefficients still suggests that changes over time remain possible, and possibly frequent. The results show that each type of problem is also related to changes in the levels of the other problem over time, albeit these reciprocal relationships are smaller than the autoregressive estimates of stability. These relationships are also slightly more pronounced between internalizing behaviors and later levels of externalizing behaviors ($\beta = -.138$ to $-.154$), than the reciprocal relationships between externalizing behaviors and later levels of internalizing behaviors ($\beta = -.099$ to $-.133$) for both groups. These reciprocal relationships are negative, suggesting that prior levels of internalizing or externalizing behaviors predict later decreases in the other type of problem once stability and time-specific correlations are taken into account. Finally, the results show that time-specific correlations between both constructs appears to be relatively small in Grades 8 and 9 ($r = .152$ to $.313$), but

become substantially larger in Grade 10 ($r = .822$ and $.952$) for both groups of adolescents.

Thus, to answer Research Question 1, these results reveal high levels of developmental stability for both constructs, as well as levels of time-specific associations between them that tend to increase over time. Furthermore, they reveal small, yet consistent, negative associations whereby levels of internalizing behaviors tend to predict decreases in externalizing behaviors over time, and levels of externalizing behaviors tend to predict decreases in internalizing behaviors over time.

Discussion

The high level of co-occurrence between internalizing and externalizing behaviors among youth is well-documented (e.g., Reitz et al. 2005). However, the temporal ordering of both types of problems remains ambiguous. In addition, the fact that youth with low levels of cognitive abilities tend to present higher levels of internalizing and externalizing behaviors than youth with average-to-high levels of cognitive abilities at any given time is also well documented (Dekker et al. 2002; Metcalfe et al. 2013; Whitaker and Read 2006). However, the extent to which knowledge regarding temporal relationships between these two types of problems generated among adolescents with average-to-high levels of cognitive abilities generalizes to youth with low levels of cognitive abilities remains essentially unknown. To address these issues, this study explored the bidirectional temporal associations between teachers' ratings of internalizing and externalizing behaviors among matched samples of adolescents with low levels of cognitive abilities and average-to-high levels of cognitive abilities extracted from a representative sample of Australian adolescents across three time waves.

Associations between Teachers' Ratings of Internalizing and Externalizing Behaviors

Our first two research questions aimed to uncover the types of relationships that would be observed between teachers' ratings of adolescents' levels of internalizing and externalizing behaviors (Research Question 1), and whether these relationships would generalize across time points (Research Question 2). Our results revealed moderately strong levels of cross-sectional associations between internalizing and externalizing behaviors that tended to increase over time, moderately high rates of stability for both types of ratings over time, and mutually suppressing relationships, also stable over time, whereby each type of behavior longitudinally predicted decreases in the other type of behavior.

Table 2 Results from Models 5 (without controls) and 12 (with controls)

	Average-to-high cognitive ability		Low cognitive ability	
	b (s.e.)	β (s.e.)	b (s.e.)	β (s.e.)
Predictive Model 5 (without controls)				
Autoregressions (stability)				
Internalizing T1→Internalizing T2 (A1)	.621* (.040)	.566 (.038)*	.621* (.040)	.654 (.058)*
Internalizing T2→Internalizing T3 (A2)	.621* (.040)	.614 (.044)*	.621* (.040)	.583 (.049)*
Externalizing T1→Externalizing T2 (B1)	.585* (.032)	.692 (.039)*	.585* (.032)	.679 (.047)*
Externalizing T2→Externalizing T3 (B2)	.585* (.032	.598 (.045)*	.585* (.032)	.647 (.045)*
Cross-lagged regressions				
Internalizing T1→Externalizing T2 (P1)	−.228* (.063)	−.138 (.040)*	−.228* (.063)	−.139 (.041)*
Internalizing T2→Externalizing T3 (P2)	−.228* (.063)	−.154 (.044)*	−.228* (.063)	−.146 (.041)*
Externalizing T1→Internalizing T2 (R1)	−.066* (.021)	−.118 (.040)*	−.066* (.021)	−.133 (.042)*
Externalizing T2→Internalizing T3 (R2)	−.066* (.021)	−.099 (.032)*	−.066* (.021)	−.108 (.033)*
Correlations				
Internalizing T1↔Externalizing T1 (C1)	.628* (.203)	.214 (.057)*	.628* (.203)	.152 (.047)*
Internalizing T2↔Externalizing T2 (C2)	.523* (.143)	.313 (.080)*	.523* (.143)	.275 (.079)*
Internalizing T3↔Externalizing T3 (C3)	1.631* (.201)	.952 (.080)*	1.631* (.201)	.822 (.091)*
Predictive Model 12 (with controls)				
Autoregressions (stability)				
Internalizing T1→Internalizing T2 (A1)	.636 (.041)*	.591 (.041)*	.636 (.041)*	.642 (.055)*
Internalizing T2→Internalizing T3 (A2)	.636 (.041)*	.627 (.045)*	.636 (.041)*	.581 (.048)*
Externalizing T1→Externalizing T2 (B1)	.593 (.032)*	.706 (.039)*	.593 (.032)*	.670 (.047)*
Externalizing T2→Externalizing T3 (B2)	.593 (.032)*	.606 (.044)*	.593 (.032)*	.641 (.045)*
Cross-lagged regressions				
Internalizing T1→Externalizing T2 (P1)	−.212 (.060)*	−.135 (.041)*	−.212 (.060)*	−.128 (.037)*
Internalizing T2→Externalizing T3 (P2)	−.212 (.060)*	−.149 (.043)*	−.212 (.060)*	−.137 (.039)*
Externalizing T1→Internalizing T2 (R1)	−.063 (.022)*	−.109 (.040)*	−.063 (.022)*	−.119 (.042)*
Externalizing T2→Internalizing T3 (R2)	−.063 (.022)*	−.090 (.032)*	−.063 (.022)*	−.096 (.033)*
Covariances/correlations				
Internalizing T1↔Externalizing T1 (C1)	.882 (.232)*	.316 (.070)*	.882 (.232)*	.325 (.147)*
Internalizing T2↔Externalizing T2 (C2)	.480 (.143)*	.293 (.083)*	.480 (.143)*	.265 (.082)*
Internalizing T3↔Externalizing T3 (C3)	1.618 (.196)*	.955 (.077)*	1.618 (.196)*	.811 (.090)*

Note As the unstandardized autoregressive coefficients (b) are fully equivalent across groups and time waves, any variations in the size of the standardized (β) coefficients are only related to variations in group- and time-specific variance estimates. s.e. = standard error of the coefficient; T1–T3 = measurement points 1 to 3; the labels A1, A2, B1, B2, P1, P2, R1, R2, C1, C2, and C3 refer to the paths labels used in Fig. 1

* $p < .01$

Increasing Levels of Cross-Sectional Associations

Looking at the pattern of relationships emerging from the final predictive model, we noted significant time-specific correlations between teachers' ratings of internalizing and externalizing behaviors. This result is interesting, especially given that the correlation between both types of problems observed in Grades 8 and 9 ($r = .152$–$.313$) is slightly lower than what could have been expected on the basis of prior research with children or adolescents (Burt et al. 2008; Gilliom and Shaw 2004; Reitz et al. 2005). A possible explanation for this lower correlation might be that the mechanisms supporting co-occurrence may slightly break down across the transition into adolescence, as youth enter puberty, get used to a new school and form a new social network (Cicchetti 2013; Eccles et al. 1993). However, as the situation in the new school stabilizes to become more familiar to the adolescents, and risk factors for both types of problems start to accumulate, then co-occurrence levels increase. It should be noted that, although the size of the time-specific correlations observed in Grade 10
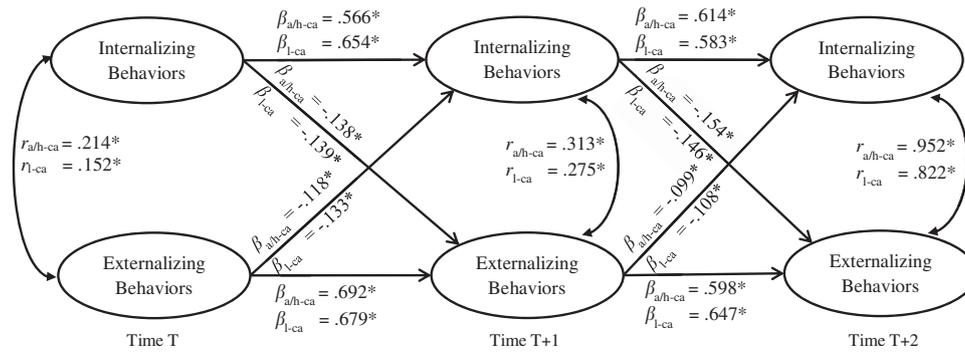
Fig. 2 Standardized Results from the Final Predictive Model (Model 5 in Table 1). *Note* As the unstandardized coefficients (*b*) are fully equivalent across groups and time waves, any variations in the size of the standardized (*β*) coefficients or correlations (*r*) are related to variations in group- and time-specific variance estimates. Coefficients accompanied by the subscript a/h-ca refer to participants with average-to-high levels of cognitive abilities, whereas those accompanied by the subscript l-ca refer to participants with low levels of cognitive abilities. *$p \leq .01$

may appear to be quite high ($r = .822$ and $.952$), this correlation represents the association between the disturbances associated with both latent constructs, that is the part of each construct remaining unexplained by the predictive paths estimated in the model. The Grade 10 correlation among these two constructs estimated as part of the preliminary measurement models ($r = .261–.531$) thus provides a more realistic estimate of the simple associations between the constructs at the end of the study. Still, these results suggest that future studies should devote attention to shared risk factors, their possible accumulation in specific individuals, and their evolution across key life transitions.

Because this study relies on teacher ratings, it can be assumed that teachers' knowledge of the internal states associated with behavioral problems as they occur in adolescents remains limited. More precisely, their assessments had to remain based strictly on observed behaviors occurring within the school context, which could possibly have led to a slight underestimation of internalizing and externalizing behaviors, and particularly for internalizing behaviors when they are hidden by externalizing behaviors. These ratings may thus have led to a slight under-estimation of the time-specific correlations between both types of behavioral problems, reinforcing the need for future research to rely on multisource assessments obtained from a wider range of informants (including students, parents and peers) in a variety of life contexts (e.g., home, peer groups, school) (for a discussion of multi-informant strategies, see De Los Reyes et al. 2015).

Longitudinal Stability

Our results also revealed that both internalizing and externalizing behaviors presented moderately high rates of longitudinal stability. Given that the results are based on teachers' ratings, which differ on a yearly basis for each specific student, the observed rates of stability cannot be assumed to reflect teachers' memory or consistency biases, through which teachers could expect a specific adolescent to display consistent behaviors across school years. Because the teacher providing the ratings changes over time, these rates of stability can be expected to reflect mainly observable behavioral stability, which is consistent with previous research showing that both internalizing and externalizing behaviors tend to present high levels of chronicity and developmental stability (e.g., Beyers and Loeber 2003; Keiley et al. 2000). Prior findings also suggest that stability may increase with age as internalizing and externalizing behaviors become more chronic (Osgood et al. 1988; Reitz et al. 2005). However, these prior studies have considered longer developmental spans as well as key life transitions, whereas the current study considers a more normative developmental period covering three subsequent secondary school years. Should rates of stability increase with age, it is thus possible that these increases may appear before (Reitz et al. 2005) or after (Osgood et al. 1988) the period covered here. Still, our results show that during adolescence, these rates of stability remained moderate, suggesting that change is possible, and that change mechanisms may be of particular interest to intervention research.

Mutually Suppressing Relationships between Internalizing and Externalizing Behaviors

Among the mechanisms underlying changes, our results revealed that these two types of problems shared negative reciprocal relationships over time, so that levels of any one type of behavioral difficulty predicted decreases in the levels of the other type of behavioral difficulty over time. In other words, our study reveals that, once the predictive cross-lagged relationships are properly controlled for the stability and co-occurrence of both types of behaviors,

internalizing and externalizing behaviors share mutually suppressing reciprocal relationships over time. These results are consistent with prior reports of negative cross-lagged relationships between internalizing and externalizing behaviors among samples of older children or adolescents (Lee and Bukowski 2012; Wiesner 2003), particularly in research based on informant reports (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016).

Although it is true that only a subset of prior studies found negative cross-lagged relationships between both types of behavioral problems, it is important to note that with two exceptions (Van der Ende et al. 2016; Van der Giessen et al. 2013), prior studies did not test the predictive equilibrium of the system. This may explain why previous results have been so inconsistent. Most previous studies including more than two time waves found relationships between internalizing and externalizing behaviors that were limited to a subset of the time intervals considered (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Klostermann et al. 2015; Masten et al. 2005; Moilanen et al. 2010; Rogosch et al. 2010; Van Lier and Koot 2010; Wiesner 2003). Without systematic tests of predictive equilibrium, it is impossible to ascertain whether differences reflect true developmental disparities, or simply random sampling variations. In the current study, we found evidence of predictive equilibrium for both the stability paths as well as for the reciprocal paths, supporting the generalizability of the results across the two time periods considered here, as well as maximizing the statistical power to detect significant relationships by the introduction of a greater level of parsimony to the model.

A particularly intriguing observation comes from the fact that prior studies of teacher's ratings conducted with younger populations all report some form of direct, or indirect, positive relationships between externalizing behaviors and later levels of internalizing behaviors, but no reciprocal relationships between internalizing behaviors and later levels of externalizing behaviors (Gooren et al. 2011; Van Lier and Koot 2010; Van Lier et al. 2012). Similarly, most studies in which negative relationships were noted between internalizing and externalizing behaviors relied on samples of adolescents or older children (Burt et al. 2008; Englund and Siebenbruner 2012; Lee and Bukowski 2012; Masten et al. 2005; Rogosch et al. 2010; Van der Ende et al. 2016; Wiesner 2003). These observations strongly suggest that the mutually suppressing relationships identified here might be specific to adolescence, and that future research should devote attention to the identification of the mechanisms at play in these developmental differences. It is important to keep in mind that the current study mainly serves to open a window into an ongoing causal chain which typically starts in childhood, and may last well into adulthood. For now, although the exact nature of the reciprocal relationships between internalizing and externalizing behaviors remains under-documented, research suggests that these relationships are much smaller than their cross-sectional associations and stability, and possibly negative.

The observation of negative, or mutually suppressing, relationships between ratings of internalizing and externalizing behaviors during the adolescent period appears inconsistent with the key theoretical models typically invoked in studies of these reciprocal relationships (e.g., Capaldi 1991 1992; Carlson and Cantwell 1980; Gold et al. 1989; Drabick et al. 2010; Moilanen et al. 2010; Oland and Shaw 2005; Wolff and Ollendick 2006), which explicitly assume these relationships to be positive and mutually reinforcing. However, the results are not inconsistent with other well-established biological or clinical theoretical models often used to describe the clinical picture of internalizing and externalizing behaviors in disconnection from one another. Thus, based on parallels drawn with the expanded fight, flight or freeze responses to threatening situations (Bracha et al. 2004; Maack et al. 2015), internalizing behaviors are often assumed to correspond to the last two types of responses, which are incompatible with the display of externalizing behaviors. Thus, although anxiety is known to result in a priming of the "fight-or-flight" response, this priming often results in avoidance behaviors, which form a critical component of its clinical picture (e.g., Zinbarg et al. 1992). Similar avoidance mechanisms are also known to be part of other forms of internalizing behaviors, such as social withdrawal (Rubin et al. 2009) and depression (Ottenbreit et al. 2014). Similarly, the Social Competition Hypothesis (Price et al. 1994, 2004; Nesse 2000) notes similarity between internalizing behaviors and primitive responses of yielding or "freezing" in the face of challenging situations, making the reliance on externalizing responses less likely.

Conversely, Snyder's (2002; Snyder and Patterson 1995; Snyder and Stoolmiller 2002) reinforcement perspective suggests that externalizing behaviors may, particularly in adolescence, represent a form of adaptive reaction to problematic environments, providing the affected individuals with a variety of negative and positive reinforcements, which may in turn limit the possibility that they will develop internalizing behaviors. The current results add to accumulating evidence (Boylan et al. 2010; Burt et al. 2008; Englund and Siebenbruner 2012; Klostermann et al. 2015; Masten et al. 2005; Moilanen et al. 2010; Rogosch et al. 2010; Van Lier and Koot 2010; Wiesner 2003) suggesting that increased levels of scientific attention should be devoted to achieving a complete understanding of the mechanisms involved in the reciprocal relationships between internalizing and externalizing behaviors. Furthermore, these results suggest a few promising theoretical

perspectives as outlined above to guide these future investigations (e.g., Kunimatsu and Marsee 2012; Price et al. 1994, 2004; Snyder 2002; Zinbarg et al. 1992) even though they have yet to be systematically applied to studies of the interrelationships between internalizing and externalizing behaviors. It is important to reinforce, however, that although we found mutually suppressing longitudinal reciprocal relationships between internalizing and externalizing behaviors during this particular developmental period, the time-specific correlations remained positive, supporting the co-occurrence of these behaviors.

It is noteworthy that some previous studies suggested that the observed relationships may differ as a function of gender (Boylan et al. 2010; Lee and Bukowski 2012; Wiesner 2003; but see Van der Ende et al. 2016). Although we showed that adding gender as a controlled variable in the analyses did not change the overall pattern of results, it was not possible to more systematically investigate whether the observed relationships differed, or not, as a function of gender due to the relatively small gender-differentiated samples. Even more generally, the analyses conducted here assume that the relationships identified in our final model applied equally to all participants. Although we did systematically investigate for the effects of one source of sample heterogeneity related to cognitive ability levels, the possibility that the observed relationships may differ as a function of other participants' characteristics has not been carefully explored. This possibility needs to be investigated more thoroughly in future research, which could potentially benefit from the adoption of a person-centered perspective aiming to directly identify sources of heterogeneity within the sample under consideration (Morin and Wang 2016; Ram and Grimm 2009). A person-centered approach might help to provide complementary perspectives on the relationships under consideration, by allowing researchers to identify subgroups of participants following distinctive longitudinal trajectories of internalizing behaviors, externalizing behaviors, and their co-occurrence.

Generalizability to Adolescents with Different Levels of Cognitive Ability

The present study systematically investigated the extent to which the longitudinal relationships between teacher ratings of internalizing and externalizing behaviors would generalize to samples of adolescents with low versus average-to-high levels of cognitive abilities. Our preliminary analyses first demonstrated that teachers' ratings of internalizing and externalizing behaviors were fully comparable (invariant) across both groups of adolescents. Given previously expressed concerns regarding the ability of youth with low levels of cognitive abilities to properly report on their internal states and behaviors (e.g., Finlay and Lyons

2001), research with this population often relies on informant reports. However, teacher reports may themselves be biased by their own assumptions regarding the way specific student characteristics (such as low levels of cognitive abilities) may influence the way constructs (such as internalizing behaviors) are manifested, or the extent to which these constructs can be observed. Our results suggest that this was not the case, and that teachers were able to provide equivalent ratings of internalizing and externalizing behaviors for adolescents with low levels of cognitive abilities, relative to their ratings of students with average-to-high levels of cognitive abilities. As further evidence of the validity of teachers' ratings, our results are in line with those from previous studies relying on alternative sources of information (Dekker et al. 2002; Whitaker and Read 2006) and support our hypothesis that adolescents with low levels of cognitive abilities tend to present higher levels of internalizing and externalizing behaviors than their peers with average-to-high levels of cognitive abilities. Our results further showed that these differences are stable in adolescence, even though internalizing behaviors slightly decreased over time for both groups (de Ruiter et al. 2007).

As a response to our final research question, our predictive model revealed that all relationships between internalizing and externalizing behaviors were fully replicated and equivalent across samples of adolescents with low versus average-to-high levels of cognitive abilities. These results are important, and tentatively suggest that current knowledge regarding the co-occurrence of internalizing and externalizing behaviors obtained within samples with average-to-high levels of cognitive abilities can be expected to generalize to youth with low levels of cognitive abilities. Although we expected that at least some of the relationships would differ as a function of the limited cognitive and social skills of youth with low levels of cognitive abilities, our results rather suggest that internalizing and externalizing behaviors tend to follow similar developmental pathways for adolescents with low levels of cognitive abilities compared to adolescents with average-to-high levels of cognitive abilities.

More precisely, a variety of theoretical perspectives (i.e., Capaldi 1991, 1992; Moilanen et al. 2010; Sameroff 1983) suggest that the more common exposure to experiences of failure in the social and academic area typically reported among adolescents with low levels of cognitive abilities should result in the observation of stronger mutually reinforcing relationships between internalizing and externalizing behaviors among this population. In contrast, because of their typically greater deficits in the cognitive and social areas, adolescents with low levels of cognitive abilities should be less likely to attain the social developmental milestones necessary for the emergence of such mutually reinforcing relationships (e.g., Oland and Shaw 2005;

Weeks et al. 2014)—leading us to expect weaker reciprocal relationships. Similarly, possible ceiling effects in the accumulation of negative experiences could limit the possibility that additional failure experiences could result in stronger relationships between internalizing and externalizing behaviors. Thus, although these various perspectives disagree with one another, they all converge in the suggestion that reciprocal relationships between internalizing and externalizing behaviors should differ (being either stronger, or weaker) among samples of adolescents with low levels of cognitive abilities when compared to their peers with average-to-high levels of cognitive abilities. Our results suggest otherwise, and reveal that the key difference appears to be related only to the quantity of symptoms (potentially due to the greater number of risk factors for youth with low levels of cognitive abilities: e.g., Rose et al. 2011; Valas 1999; Wehmeyer 2005), rather than to more qualitative differences in the way these symptoms relate to one another. Still, our results also suggest the need for additional theoretical developments in this area. Indeed, a common characteristic of these prior theoretical propositions is that they all assumed that relationships between internalizing and externalizing behaviors would be mutually reinforcing. However, by revealing mutually suppressing relationships between internalizing and externalizing behaviors, our results suggest that these various theories may not be able to properly explain reciprocal relationships between internalizing and externalizing behaviors as they occur in adolescence. Understanding why these relationships turned out to be similar, rather than different, among these two populations should be anchored in a better understanding of why these relationships are mutually suppressing in the first place. Although we proposed a variety of tentative explanations for these results, future research will be clearly needed to test, revise, and improve on these explanations.

Finally, it is important to keep in mind that the type of educational placement experienced by adolescents with low levels of cognitive abilities is an important contextual factor impacting upon their development (e.g., Tracey et al. 2003). The Wollongong Youth Study, which provided the sample for this study, did not collect information about adolescents' type of educational placement and did not include specialized establishments, which limits our ability to consider the role of this factor and populations experiencing the lowest levels of cognitive functioning. Similarly, the current study adopted a non-categorical approach focusing on student's level of cognitive functioning rather than focusing on the etiology of their cognitive difficulties (e.g., Autism, Down syndrome). Still, information about specific developmental and learning disabilities, had such information been available, might have provided us with a more nuanced understanding of the mechanisms involved in the development of internalizing and externalizing behaviors for these

adolescents. In addition, cognitive abilities were assessed through a standardized test of literacy and numeracy. Despite evidence, previously presented, that this test provides a highly reliable proxy for IQ, they remain, at best, an indirect measure of IQ and cognitive functioning. As such, future research would do well to more carefully consider the impact of school placements, diagnoses of developmental disabilities and cognitive levels as assessed through formal standardized IQ tests on the current results.

Conclusion

Internalizing and externalizing behaviors are frequent among adolescents, accompanied by a variety of negative developmental consequences, and often co-occurring. So far, this co-occurrence has been assumed to reflect the action of common determinants, but also the presence of mutually reinforcing relationships between these two types of behavioral problems. For preventive purposes, the presence of such mutually reinforcing relationships would suggest the possibility of preventing, or at least reducing, the risk for both types of behavioral problems through interventions targeting only one type of behavioral problem. However, prior research has led to ambiguous results regarding the nature of the reciprocal relationships found among these two types of behavioral problems, particularly during adolescence where accumulating evidence suggests that these relationships could be mutually suppressing, rather than reinforcing. In providing further evidence that these relationships were indeed mutually suppressing, the present study calls for improved theoretical developments in this area. More importantly, it suggests that preventive interventions cannot simply assume that reducing the risk for one type of behavioral problem would directly translate to a reduction in the risk of the other behavioral problem. Rather, it suggests the importance of developing joint programs targeting both types of behavioral problems.

Another key contribution of the present study is the observation that the relationships fully generalized to adolescents with low levels of cognitive abilities, which are known to present a higher level of risk for both types of behavioral problems than their peers with average-to-high levels of cognitive abilities (Dekker et al. 2002; Metcalfe et al. 2013; Whitaker and Read 2006). This result is particularly important given the relative rarity of research focusing on samples of youth with low levels of cognitive abilities, which can be partly explained by the challenges posed by conducting research with this population (i.e., difficulty of access, complexity of obtaining reliable measures, smaller samples, etc.). Possibly because of these challenges, developmental research on youth with low levels of cognitive abilities has often been conducted in

relative disconnection from developmental research conducted with more normative populations, leading often to the development of interventions that are specific to youth with low levels of cognitive abilities. By providing evidence that the pattern of relationships between internalizing and externalizing behaviors generalizes to both samples, the present research suggests that the more extensive knowledge base accumulated with normative populations may generalize to samples of adolescents with low levels of cognitive abilities—as least in regards to the constructs considered in the present study. In doing so, these results also suggest the possibility to develop preventive interventions that would equally apply to all adolescents, pending further research aiming to better document the extent to which the mechanisms involved in the development of both types of behavioral problems also generalize to both populations.

Author's Contributions AJSM, AKA, and CM conceived the study together, realized the analyses, and wrote the article, with a slightly greater involvement on the part of AJSM and AKA during a visiting professor stay at the German Institute for International Educational Research; JC collected the data set and contributed to the interpretations and writing of the article; DT and RG provided expertise related to youth with low levels of cognitive abilities, helped to integrate this information to the article, and contributed to the interpretations and writing of the article; PDP provided statistical expertise to the paper by conducting the matching process, and helping in the design and writing of the statistical sections. All authors read and approved the final version of article.

Compliance with Ethical Standards

Appendix: Preliminary Measurement Models

We estimated Confirmatory Factor Analytic (CFA) models underlying teachers' ratings of internalizing and externalizing behaviors across both groups of adolescents and each of the three time waves. This multiple-group longitudinal CFA included a total of six factors (internalizing and externalizing behaviors × three time waves) in both groups of adolescents. These factors were specified as congeneric, with each indicator allowed to load on a single factor. The models also included a priori correlated uniquenesses between matching indicators utilized at the different time points to reflect the fact that the indicators' unique variance emerges in part from shared sources of influences over time (Mitchison et al. 2015). A critical assumption of multiple group longitudinal models is that the measurement model underlying the constructs is invariant (i.e., equivalent) across groups and time (Millsap 2011). To verify this assumption, we tested the measurement invariance of the complete CFA model across groups and time points in the following sequence (Millsap 2011; Morin et al. 2011): (i) configural invariance (the same measurement model with the same number of factors defined by the same items), (ii) weak invariance (invariant loadings); (iii) strong invariance (invariant loadings and thresholds); (iv) strict invariance (invariant factor loadings, item thresholds, and uniquenesses). Model fit assessment followed the procedures described in the main manuscript.

The goodness-of-fit from these measurement models is reported in Table 3. These results show that the baseline model of configural invariance (Model 1) provides an acceptable level of fit to the data according to CFI and TLI ≥ .900 and an excellent fit to the data according to RMSEA ≤ .060. The results further show that adding measurement invariance constraints across groups and time on the factors loadings (Model 2), thresholds (Model 3), and uniquenesses (Model 4) never resulted in a decrease of fit greater than the recommended guidelines. Fit indices including a correction for parsimony (TLI, RMSEA) even systematically improved with these added constraints, thus supporting the strict measurement invariance of this model across groups and time waves. Parameters estimates from this strictly invariant CFA model are reported in Tables 4 and 5, together with estimates of composite reliability. Composite reliability was calculated using McDonald's (1970) omega ($\omega = (\Sigma|\lambda i|)^2/([\Sigma|\lambda i|]^2 + \Sigma\delta ii)$ where λi are the factor loadings and δii, the error variances), which is similar to alpha but takes into account the strength of association between items and constructs as well as item-specific measurement errors. Supporting the strength of the measurement model, all factor loadings and estimates of composite reliability were relatively high and satisfactory across groups and time points: (a) internalizing behaviors ($\lambda = .610-.877$; $\omega = .904-.937$); (b) externalizing behaviors ($\lambda = .660-.933$; $\omega = .959-.974$).

Table 3 Goodness-of-fit indices for the preliminary measurement models

		χ^2	df	CFI	TLI	RMSEA	90 % CI	CM	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA
Measurement models													
1	Configural invariance	4807.638*	3270	.914	.907	.037	[.035; .040]	–	–	–	–	–	–
2	Weak invariance (loadings) across time and groups	4891.799*	3360	.915	.910	.037	[.035; .039]	1	110.169	90	+.001	+.003	.000
3	Strong invariance (loadings, thresholds) across time and groups	4996.784*	3500	.917	.916	.036	[.033; .038]	2	190.082*	140	+.002	+.006	−.001
4	Strict invariance (loadings, thresholds, uniquenesses) across time and groups	5084.372*	3600	.917	.919	.035	[.033; .037]	3	220.440*	100	.000	+.003	−.001

χ^2 WLSMV chi square test of exact fit, *df* degrees of freedom, *CFI* comparative fit index, *TLI* Tucker-Lewis index, *RMSEA* root mean square error of approximation, *CI* confidence interval, *CM* nested comparison model (typically the previous model in the sequence), $\Delta\chi^2$ chi-square difference test for WLSMV estimation calculated using the Mplus DIFFTEST function, Δ Change relative to the comparison model (CM)

*$p \leq .01$

Table 4 Standardized factor loadings (λ) and uniquenesses (δ) from the strictly invariant multiple-group longitudinal measurement model

	Time 1				Time 2				Time 3			
	A/H-CA		L-CA		A/H-CA		L-CA		A/H-CA		L-CA	
	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ
Internalizing behaviors												
Indicator 1	.781	.390	.824	.322	.845	.286	.817	.333	.820	.327	.816	.333
Indicator 2	.706	.502	.756	.428	.783	.388	.748	.440	.752	.434	.748	.441
Indicator 3	.610	.629	.666	.557	.696	.515	.657	.569	.662	.562	.656	.570
Indicator 4	.783	.388	.825	.320	.846	.284	.818	.330	.822	.325	.818	.331
Indicator 5	.793	.371	.834	.304	.854	.270	.828	.315	.831	.309	.827	.316
Indicator 6	.822	.324	.859	.263	.877	.232	.853	.272	.856	.267	.853	.273
Indicator 7	.796	.366	.837	.300	.857	.266	.831	.310	.834	.305	.830	.311
ω	.904		.926		.937		.923		.925		.923	
Externalizing behaviors												
Indicator 1	.933	.129	.932	.131	.914	.164	.916	.161	.896	.197	.916	.161
Indicator 2	.872	.240	.870	.244	.839	.295	.842	.291	.810	.344	.842	.291
Indicator 3	.768	.411	.765	.415	.721	.480	.725	.475	.681	.536	.725	.475
Indicator 4	.749	.439	.746	.443	.701	.509	.705	.503	.660	.564	.705	.503
Indicator 5	.898	.194	.896	.197	.870	.242	.873	.238	.845	.286	.873	.238
Indicator 6	.897	.196	.895	.199	.869	.244	.872	.240	.844	.288	.872	.240
Indicator 7	.919	.155	.918	.157	.897	.196	.899	.192	.876	.233	.899	.192
Indicator 8	.881	.225	.879	.228	.850	.277	.853	.273	.822	.324	.853	.273
Indicator 9	.925	.144	.924	.146	.904	.182	.906	.179	.884	.218	.906	.179
Indicator 10	.793	.372	.790	.376	.749	.440	.752	.434	.710	.495	.752	.434
Indicator 11	.873	.238	.871	.241	.841	.293	.844	.288	.812	.341	.844	.288
Indicator 12	.922	.150	.921	.152	.901	.189	.903	.185	.880	.226	.902	.186
Indicator 13	.770	.407	.768	.411	.724	.476	.728	.470	.684	.532	.728	.470
ω	.974		.974		.967		.968		.959		.968	

ω McDonald's omega coefficient of composite reliability, *L-CA* group of adolescents presenting low levels of cognitive abilities, *A/H-CA* group of adolescents presenting average-to-high levels of cognitive abilities

All loadings are significant *(p < .05)*

Table 5 Latent and observed correlations from the strictly invariant multiple-group longitudinal measurement model

	Age	Gender	Minority	SES	IB T1	IB T2	IB T3	EB T1	EB T2	EB T3
Age		−.164	.014	.020	−.017	−.075	−.044	.084	.015	.066
Gender	−.165		−.032	.114	.128*	.110*	.075	−.313*	−.285*	−.297*
Minority	.027	−.043		.029	.099	.042	.016	.075	−.028	−.075
SES	−.018	.042	−.040		−.136*	−.084	−.176*	−.074	−.069	−.096
IB T1	.146	.033	−.043	.101		.505*	.500*	.226*	−.012	.029
IB T2	.145	.197*	.106	.120	.483*		.491*	.039	.111*	−.044
IB T3	.202*	.112	.038	−.166	.466*	.500*		.063	.031	.531*
EB T1	−.075	−.259*	−.127	.120	.103	−.064	−.119		.562*	.507*
EB T2	−.135	−.297*	−.085	.012	−.133	−.079	−.264	.639*		.452*
EB T3	.106	−.273*	−.051	−.054	−.116	−.086	.261*	.549*	.626*	

Note Results from the group of adolescents presenting low levels of cognitive abilities reported under the diagonal; results from the group of adolescents presenting average-to-high levels of cognitive abilities reported above the diagonal

SES socio-economic status, *IB* internalizing behaviors, *EB* externalizing behaviors, *T1–T3* measurement points 1–3

*$p < .05$

References

Achenbach, T. M. (2009). *The Achenbach System of Empirically Based Assessment (ASEBA)*. Burlington, VT: University of Vermont.

Akse, J., Hale, B., Engels, R., Raaijmakers, Q., & Meeus, W. (2007). Co-occurrence of depression and delinquency in personality types. *European Journal of Personality*, 21, 235–256.

Asparouhov, T., & Muthén, B.O. (2010). Weighted Least Square estimation with missing data. http://www.statmodel.com/download/GstrucMissingRevision.pdf.

Australian Bureau of Statistics. (2005). Yearbook Australia: Population. Canberra, Australia: Government Printer (Document 1301.0).

Australian Bureau of Statistics (2006). ANZSCO - Australian and New Zealand Standard Classification of Occupations, 1st edn. (Document. 1220.0).

Australian Bureau of Statistics. (2012). *Australian Social Trends*. Canberra, Australia: Government Printer. (Document 4102.0).

Australian Curriculum, Assessment & Reporting Authority. (2015). *NAPLAN Achievement in reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2015*. Sydney: ACARA.

Beyers, J. M., & Loeber, R. (2003). Untangling developmental relations between depressed mood and delinquency in male adolescents. *Journal of Abnormal Child Psychology*, 31, 247–266.

Boylan, K., Georgiades, K., & Szatmari, P. (2010). The longitudinal association between oppositional and depressive symptoms across childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49, 152–161.

Bracha, H. S., Williams, A. E., Ralston, T. C., Bracha, A. S., & Matsukawa, J. M. (2004). Does "fight or flight" need updating? *Psychosomatics*, 45, 448–449.

Brentsilver, M., Negriff, S., Mennen, F. E., & Trickett, P. K. (2011). Longitudinal relations between depressive symptoms and externalizing behavior in adolescence: Moderating effects of maltreatment and gender. *Journal of Clinical Child & Adolescent Psychology*, 40, 607–617.

Buist, K. L., Dekovic, M., Meeus, W., & van Aken, M. A. G. (2004). The reciprocal relationship between early adolescent attachment and internalizing and externalizing problem behavior. *Journal of Adolescence*, 27, 251–266.

Burt, K. B., Obradović, J., Long, J., & Masten, A. (2008). The interplay of social competence and psychopathology over 20 years. *Child Development*, 79, 359–374.

Capaldi, D. M. (1991). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: I. Family factors and general adjustment at grade 6. *Development & Psychopathology*, 3, 277–300.

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up Grade 8. *Development & Psychopathology*, 4, 125–144.

Carlson, G. A., & Cantwell, D. P. (1980). Unmasking masked depression in children and adolescents. *American Journal of Psychiatry*, 137, 445–449.

Caron, C., & Rutter, M. (1991). Co-morbidity in child psychopathology: Concepts, issues, and research strategies. *Journal of Child Psychology & Psychiatry*, 132, 1063–1080.

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

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.

Ciarrochi, J., Heaven, P. C. L., & Davies, F. (2007). The impact of hope, self-esteem, and attributional style on adolescents' school grades and emotional well-being: A longitudinal study. *Journal of Research in Personality*, 41, 1161–1178.

Ciarrochi, J., Heaven, P. C. L., & Skinner, T. (2012). Cognitive ability and health-related behaviors during adolescence: A prospective study across five years. *Intelligence*, 40, 317–324.

Cicchetti, D. (2013). Developmental psychopathology. In P. Zelazo (Ed.), *Oxford Handbook of Developmental Psychology* 2, (pp. 455–480). New York: Oxford University.

Connell, A. M., & Dishion, T. J. (2006). The contribution of peers to monthly variation in adolescent depressed mood: A short-term longitudinal study with time varying predictors. *Development & Psychopathology*, 18, 139–154.

Craven, R. G., Morin, A. J. S., Tracey, D., Parker, P. D., & Zhong, H. F. (2015). *Inclusive Education for Students with Intellectual Disabilities*. Charlotte, NC: IAP.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 105–136). Washington, DC: American Psychological Association.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data. *Journal of Abnormal Psychology, 112*, 558–577.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

Dekker, M. C., Koot, H. M., Van der Ende, J., & Verhulst, F. C. (2002). Emotional and behavioral problems in children and adolescents with and without intellectual disability. *Journal of Child Psychology & Psychiatry, 43*, 1087–1098.

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

de Ruiter, K. P., Dekker, M. C., Verhulst, F. C., & Koot, H. M. (2007). Developmental course of psychopathology in youths with and without intellectual disabilities. *Journal of Child Psychology and Psychiatry, 48*, 498–507.

Drabick, D. A. G., Olendick, T. H., & Bubier, J. L. (2010). Co-occurrence of ODD and anxiety: Shared risk processes and evidence for a dual-pathway model. *Clinical Psychology, 17*, 307–318.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*, 90–101.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Englund, M. M., & Siebenbruner, J. (2012). Developmental pathways linking externalizing symptoms, internalizing symptoms, and academic competence to adolescent substance use. *Journal of Adolescence, 35*, 1123–1140.

Fanti, K. A., & Henrich, C. C. (2010). Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology, 46*, 1159–1175.

Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds), *Structural Equation Modeling: A Second Course*. 2nd edn. (pp. 439–492). Greenwich, CO: IAP.

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

Frey, M. C., & Detterman, D. (2004). Scholastic assessment or g? *Psychological Science, 15*, 373–378.

Gilliom, M., & Shaw, D. S. (2004). Co-development of externalizing and internalizing problems in early childhood. *Development & Psychopathology, 16*, 313–333.

Gold, M., Mattlin, J., & Osgood, D. W. (1989). Background characteristics and responses to treatment of two types of institutionalized delinquent boys. *Criminal Justice and Behavior, 16*, 5–33.

Gooren, E. M., van Lier, P. A., Stegge, H., Terwogt, M. M., & Koot, H. M. (2011). The development of conduct problems and depressive symptoms in early elementary school children: the role of peer rejection. *Journal of Clinical Child & Adolescent Psychology, 40*, 245–253.

Hayes, L., & Ciarrochi, J. (2015). *The thriving adolescent: Using acceptance and commitment therapy and positive psychology to help young people manage emotions, achieve goals, and build positive relationships*. Oakland, CA: New Harbinger.

Heaven, P. C. L., & Ciarrochi, J. (2008). Parental styles, conscientiousness, and academic performance: A three-wave longitudinal study. *Personality & Social Psychology Bulletin, 34*, 451–461.

Heaven, P. C. L., & Ciarrochi, J. (2012). When IQ is not everything: Intelligence, personality and academic performance at school. *Personality & Individual Differences, 53*, 518–522.

Heaven, P. C. L., Ciarrochi, J., & Leeson, P. (2011). Cognitive ability, right-wing authoritarianism, and social dominance orientation: A five-year longitudinal study. *Intelligence, 39*, 15–21.

Heaven, P. C. L., Leeson, P., & Ciarrochi, J. (2009). Personality development at school: Assessing a reciprocal influence model of teachers' evaluations and students' personality. *Journal of Research in Personality, 43*, 815–821.

Helps, S. (2015). Psychopathology: Anxiety, Depression and Schizophrenia. In J. L. Matson, & M. L. Matson (Eds.), *Comorbid conditions in individuals with intellectual disabilities* (pp. 85–107). Cham, Switzerland: Springer.

Herrenkohl, T. I., Kosterman, R., Mason, W. A., Hawkins, J. D., McCarty, C. A., & McCauley, E. (2010). Effects of childhood conduct problems and family adversity on health, health behaviors, and service use in early adulthood: Tests of developmental pathways involving adolescent risk taking and depression. *Development & Psychopathology, 22*, 655–665.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software, 42*, 1–28.

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

Jackson, M., Jonsson, J. O., & Rudolphi, F. (2012). Ethnic inequality in choice-driven education systems: A longitudinal study of performance and choice in England and Sweden. *Sociology of Education, 85*, 158–178.

Jackson, Y., Frick, P., & Dravage-Bush, J. (2000). Perceptions of control in children with externalizing and mixed behavior disorders. *Child Psychiatry & Human Development, 31*, 43–58.

Joiner, Jr., T. E., & Coyne, J. C. (1999). *The interactional nature of depression: Advances in interpersonal approaches*. Washington, DC: American Psychological Association.

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology, 28*, 161–179.

Kiddle, H., & Dagnan, D. (2011). Vulnerability to depression in adolescents with intellectual disabilities. *Advances in Mental Health and Intellectual Disabilities, 5*, 3–8.

Klostermann, S., Connel, A., & Stormshak, E.A. (2015, in press). Gender differences in the developmental links between conduct problems and depression across early adolescence. Journal of Research on Adolescence. Advance online publication. doi: 10.1111/jora.12170.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2*, 111–133.

Kunimatsu, M. M., & Marsee, M. A. (2012). Examining the presence of anxiety in aggressive individuals: The illuminating role of fight-or-flight mechanisms. *Child Youth Care Forum, 41*, 247–258.

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling, 18*, 649–662.

Lee, E. J., & Bukowski, W. M. (2012). Co-development of internalizing and externalizing problem behaviors: Causal direction and common vulnerability. *Journal of Adolescence, 35*, 713–729.

Lee, E. J., & Stone, S. I. (2012). Co-occurring internalizing and externalizing behavioral problems: The mediating effect of negative self-concept. *Journal of Youth & Adolescence, 41*, 717–731.

Maack, D. J., Buchanan, E., & Young, J. (2015). Development and psychometric investigation of an inventory to assess fight, flight, and freeze tendencies: The Fight, Flight, and Freeze Questionnaire. *Cognitive Behaviour Therapy, 44*, 117–127.

Marsh, H. W., Hau, K. -T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares, & J. McArdle (Eds.), *Contemporary psychometrics. A Festschrift for Roderick P. McDonald*. Mahwah NJ: Erlbaum.

Marsh, H. W., Hau, K. -T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling, 11*, 320–341.

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

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradovic, J., Riley, J., Boelcke-Stennes, K., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology, 41*, 733–746.

Mathiesen, K. S., Sanson, A., Stoolmiller, M., & Karevold, E. (2009). The nature and predictors of undercontrolled and internalizing problem trajectories across early childhood. *Journal of Abnormal Child Psychology, 37*, 209–222.

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

Measelle, J. R., Stice, E., & Hogansen, J. M. (2006). Developmental trajectories of co-occurring depressive, eating, antisocial, and substance abuse problems in female adolescents. *Journal of Abnormal Psychology, 115*, 524–538.

Metcalfe, L. A., Harvey, E. A., & Laws, H. B. (2013). The longitudinal relation between academic/cognitive skills and externalizing behavior problems in preschool children. *Journal of Educational Psychology, 105*, 881.

Miers, A. C., Blöte, A. W., Heyne, D. A., & Westenberg, P. M. (2014). Developmental pathways of social avoidance across adolescence: The role of social anxiety and negative cognition. *Journal of Anxiety Disorders, 28*, 787–794.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Mitchison, D., Morin, A. J. S., Mond, J., Slewa-Younan, S., & Hay, P. (2015). The bidirectional relationship between quality of life and eating disorder symptoms: A 9-year community-based study of Australian women. *PLOS One, 10*, e0120591.

Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development & Psychopathology, 22*, 635–653.

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 Non-Clinical Adults. *Epidemiology and Public Health/Revue d'Épidémiologie et de Santé Publique, 59*, 327–340.

Morin, A. J. S., & Wang, J. C. K. (2016). A gentle introduction to mixture modeling using physical fitness data. In N. Ntoumanis, & N. Myers (Eds.), *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists* (pp. 183–210). London, UK: Wiley.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Nesse, R. M. (2000). Is depression an adaptation? *Archives of General Psychiatry, 57*, 14–20.

Newman, D., Moffitt, T., Caspi, A., & Silva, P. (1998). Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology, 107*, 305–311.

Oland, A. A., & Shaw, D. S. (2005). Pure versus co-occurring externalizing and internalizing symptoms in children: The potential role of socio-developmental milestones. *Clinical Child & Family Psychology Review, 8*, 247–270.

Organisation for European Co-Operation and Development (OECD). (1999). *Inclusive education at work: students with disabilities in mainstream schools*. Paris: OECD.

Osgood, D. W., Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1988). The generality of deviance in late adolescence and early adulthood. *American Sociological Review, 53*, 81–93.

Ottenbreit, N. S., Dobson, K. S., & Quigley, L. (2014). An examination of avoidance in major depression in comparison to social anxiety disorder. *Behaviour Research & Therapy, 56*, 82–90.

Overbeek, G., Vollebergh, W., Meeus, W., Engels, R., & Luijpers, E. (2001). Course, co-occurrence, and longitudinal associations of emotional disturbance and delinquency from adolescence to young adulthood: A six-year three-wave study. *Journal of Youth & Adolescence, 30*, 401–426.

Price, J. S., Gardner, R., & Erickson, M. (2004). Can depression, anxiety, and somatization be understood as appeasement displays?. *Journal of Affective Disorders, 79*, 1–11.

Price, J., Sloman, L., Gardner, Jr., R., Gilbert, P., & Rohde, P. (1994). The social competition hypothesis of depression. *British Journal of Psychiatry, 164*, 309–315.

Pulkkinen, L., Kaprio, J., & Rose, R. J. (1999). Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: The multidimensional peer nomination inventory. *Twin Research, 2*, 274–285.

Quiroga, C. V., Janosz, M., Bisset, S., & Morin, A. J. S. (2013). Early adolescent depression symptoms and school dropout: Mediating processes involving self-reported academic competence and achievement. *Journal of Educational Psychology, 105*, 552–560.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*, 565–576.

Reitz, E., Deković, M., & Meijer, A. (2005). The structure and stability of externalizing and internalizing problems during early adolescence. *Journal of Youth & Adolescence, 34*, 577–588.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A cmparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373.

Rogosch, F. A., Oshri, A., & Cicchetti, D. (2010). From child maltreatment to adolescent cannabis use and dependence: A developmental cascade model. *Development & Psychopathology, 22*, 883–897.

Rose, C. A., Monda-Amaya, L. E., & Espelage, D. L. (2011). Bullying perpetration and victimization in special education: A review of the literature. *Remedial and Special Education, 32*, 114–130.

Rubin, K. H., Coplan, R. J., & Bowker, J. C. (2009). Social withdrawal in childhood. *Annual Review of Psychology, 60*, 141–171.

Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In W. Kessen (Ed.), *Handbook of child psychology* 1, (pp. 237–294). New York: Wiley.

Snyder, J. (2002). Reinforcement and coercion mechanisms in the development of antisocial behavior: Peer relationships. In J. B.

Reid, G. R. Patterson, & J. Snyder (Eds.), *Antisocial behavior in children and adolescents* (pp. 101–122). Washington, DC: American Psychological Association.

Snyder, J., & Patterson, G. R. (1995). Individual differences in social aggression: A test of a reinforcement model of socialization in the natural environment. *Behavior Therapy*, 26, 371–391.

Snyder, J., & Stoolmiller, M. (2002). Reinforcement and coercion mechanisms in the development of antisocial behavior: the family. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.), *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention* (pp. 65–100). Washington, DC: American Psychological Association.

Stone, L. L., Otten, R., Engels, R. C. M. E., Kuijpers, R. C. W. M., & Janssens, J. M. A. M. (2015). Relations between internalizing and externalizing problems in early childhood. *Child Youth Care Forum*, 44, 635–653.

Thoemmes, F., & Kim, E. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, 46, 90–118.

Tracey, D., Marsh, H. W., & Craven, R. G. (2003). Self-concepts of preadolescents with mild intellectual disabilities. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), International advances in self research (pp. 203–229). Greenwich, CT: Information Age.

Valas, H. (1999). Students with learning disabilities and low-achieving students: Peer acceptance, loneliness, self-esteem, and depression. *Social Psychology of Education*, 3, 173–192.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.

Van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2016). The bidirectional pathways between internalizing and externalizing problems and academic performance from 6 to 18 years. *Development and Psychopathology*, 28, 855–867.

Van der Giessen, D., Branje, S., Overbeek, G., Frijns, T., van Lier, P. A. C., Koot, H. M., & Meeus, W. (2013). Co-occurrence of aggressive behavior and depressive symptoms in early adolescence: A longitudinal multi-informant study. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 63, 193–201.

Van Lier, P. A. C., & Koot, H. M. (2010). Developmental cascades of peer relations and symptoms of externalizing and internalizing problems from kindergarten to fourth-grade elementary school. *Development & Psychopathology*, 22, 569–582.

Van Lier, P. A., Vitaro, F., Barker, E. D., Brendgen, M., Tremblay, R. E., & Boivin, M. (2012). Peer victimization, poor academic achievement, and the link between childhood externalizing and internalizing problems. *Child Development*, 83, 1775–1788.

Vieno, A., Kiesner, J., Pastore, M., & Santinello, M. (2008). Antisocial behavior and depressive symptoms: Longitudinal and concurrent relations. *Adolescence*, 43, 649–660.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., & Maelender, A. (2003). *Wechsler intelligence scale for children: Fourth edition*. San Antonio, TX: Pearson.

Weeks, M., Wild, T. C., Ploubidis, G. B., Naicker, K., Cairney, J., North, C. R., & Colman, I. (2014). Childhood cognitive ability and its relationship with anxiety and depression in adolescence. *Journal of Affective Disorders*, 152, 139–145.

Wehmeyer, M. L. (2005). Self-determination and individuals with severe disabilities. *Research & Practice for Persons with Severe Disabilities*, 30, 113–120.

Westwood, P., & Graham, L. (2000). How many children with special needs in regular classes? Official predictions VS teachers' perceptions in South Australia and New South Wales. *Australian Journal of Learning Disabilities*, 5, 24–35.

Wiesner, M. (2003). A longitudinal latent variable analysis of reciprocal relations between depressive symptoms and delinquency during adolescence. *Journal of Abnormal Psychology*, 112, 633–645.

Whitaker, S., & Read, S. (2006). The prevalence of psychiatric disorders among people with intellectual disabilities: An analysis of the literature. *Journal of Applied Research in Intellectual Disabilities*, 19, 330–345.

Wolff, J. C., & Ollendick, T. H. (2006). The comorbidity of conduct problems and depression in childhood and adolescence. *Clinical Child & Family Psychology Review*, 9, 201–220.

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

Zinbarg, R. E., Barlow, D. H., Brown, T. A., & Hertz, R. M. (1992). Cognitive behavioral approaches to the nature and treatment of anxiety disorders. *Annual Review of Psychology*, 43, 235–267.

Alexandre J. S. Morin is a full professor at the Institute for Positive Psychology and Education of the Australian Catholic University (New South Wales, Australia). He received his Ph.D. in 2005 from the Université de Montréal (Québec, Canada). His major research interests include the exploration of the social determinants of psychological well-being and psychopathologies at various life stages and in various settings, such as schools and organizations.

A. Katrin Arens is a postdoctoral researcher in the Department of Education and Human Development and Centre for Research on Individual Development and Adaptive Education of the German Institute for International Educational Research (Germany). She received her Ph.D. in 2011 from Universität Göttingen (Germany). Her major research interests include the structure, development, and formation of students' multidimensional self-concept and academic motivation.

Christophe Maïano is an associate professor in the Département de Psychoéducation et de Psychologie of the Université du Québec en Outaouais (Québec, Canada). He received his Ph.D. in 2003 from the Université de Montpellier I (France). His major research interests include psychosocial interventions, adapted physical activity interventions, and health prevention and education interventions, all aimed at improving the physical and psychological well-being of individuals with an intellectual disability (ID), physical health problems, or mental health disorders.

Joseph Ciarrochi is a full professor at the Institute for Positive Psychology and Education of the Australian Catholic University (New South Wales, Australia). He received his Ph.D. in 1996 from the University of Pittsburgh (PA, USA). His major research interests include the identification of character strengths that promote social, emotional, physical well-being and performance, and contextual behavioral science.

Danielle Tracey is a senior lecturer at the Western Sydney University (New South Wales, Australia). She received her Ph.D. in 2002 from the University of Western Sydney (New South Wales, Australia). Her major research interests include the development and management of community and school-based programs for children, young people,

and families, particularly in the area of inclusive education for youth with intellectual disabilities.

Philip D. Parker is a senior lecturer and an Australian Research Council funded researcher at the Institute for Positive Psychology and Education at the Australian Catholic University (New South Wales, Australia). He received his Ph.D. in 2010 from the University of Sydney (New South Wales, Australia). His major research interests include educational inequality, developmental transitions, and educational attainment.

Rhonda G. Craven is a full professor and institute director at the Institute for Positive Psychology and Education of the Australian Catholic University (New South Wales, Australia). She received her Ph.D. in 1996 from the University of Sydney (New South Wales, Australia). Her major research interests include the structure of self-concept and key psychosocial drivers of well-being and performance, with specialised expertise in Indigenous education.