Can Teachers' Need-Supportive Practices Moderate the Big-Fish-Little-Pond Effect? A Quasi-Experimental Study with Elementary School Children

William Gilbert[1], Frédéric Guay[2], and Alexandre J.S. Morin[3]

[1] School of Psychology, Université Laval, Canada

[2] Department of Educational Fundamentals and Practices, Université Laval, Canada

[3]Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada

Author Notes

The authors have no conflict of interest to report.
This research was supported by the Canada Research Chair Program (201844), and the Fonds de Recherche sur la Société et la Culture (FQRSC) (13915).
Correspondence concerning this article should be addressed to Frédéric Guay, Department of Educational Fundamentals and Practices, Université Laval, 2320 rue des Bibliothèques, Quebec City, Quebec, G1V 0A6, Canada. E-mail: frederic.Guay@fse.ulaval.ca

This is the prepublication version of the following manuscript:

Gilbert, W., Guay, F., & Morin, A.J.S. (in press). Can teachers' need-supportive practices moderate the Big-Fish-Little-Pond Effect? A quasi-experimental study with elementary school children, *Contemporary Educational Psychology*. Early view doi: https://doi.org/10.1016/j.cedpsych.2022.102060

Abstract

The big-fish-little-pond effect (BFLPE) describes a multilevel phenomenon via which class-average levels of achievement act as a negative predictor of students' academic self-concept (ASC), although their individual achievement level acts as a positive predictor of their ASC. In this quasi-experimental pretest-posttest study, we posited that a professional development program for teachers based on self-determination theory would moderate the BFLPE on French writing ASC among elementary school students. To test our hypothesis, we relied on three groups of students (n = 189 to 394) who were exposed to teachers who participated in the program at different moments, and one group of students (n = 190) who were exposed to teachers who did not participate in any professional development program. Doubly latent multilevel analyses were used to test the moderation hypothesis. At posttest, we observed a significant BFLPE for the control students while no significant BFLPE was observed for the other groups of students. These results suggest that this professional development program was able to protect students against the BFLPE. Promoting a learning context supportive of students' psychological needs thus seem able to reduce the salience of social comparison processes involving the achievement levels of the whole class by reinforcing the focus on students' own learning experiences. These results support previous research suggesting that positive teacher-student relationships might moderate the BFLPE. Finally, our findings provide insights regarding the pedagogical practices that can attenuate the negative effect of class-average levels of achievement on students' writing ASC.

***Keywords*:** Elementary grades; Self-concept; BFLPE; Self-determination theory; Basic psychological needs; Social comparison.

Students are confronted with social comparison information in a variety of domains (e.g., physical appearance, popularity), including academic performance. School grades are indeed among the most salient sources of social comparison in the classroom (Dijkstra et al., 2008). This is partly due to the nature of the school system, which is based on the premise that students must succeed at specific tasks to be promoted to the next grade level. This reward system based on performance is endorsed not only by teachers, but also by parents who often put pressure on their children to motivate them to succeed at school. This evaluative atmosphere evokes in children a strong interest in comparing themselves with others to find out how competent they are in various school subjects (Buunk et al., 2005).

Marsh (1984, 1987; Marsh & Parker, 1984) proposed the "big-fish-little-pond effect" (BFLPE) to describe one key social comparison process taking place in schools or classrooms. More precisely, the BFLPE relies on the assumption that students compare their individual achievement with the average performance of other students attending the same school and/or classroom. The information acquired through this social comparison process is in turn used to inform the development of their academic self-concept (ASC), which refers to students' evaluative self-perceptions of their abilities in different school subjects (Marsh & Craven, 1997; Shavelson et al., 1976). The BFLPE thus proposes that students who attend schools or classroom composed of more able peers (characterized by higher average levels of achievement) should make less favorable social comparisons, and thus form a more negative ASC than their equally able counterparts educated in mixed- or low-ability schools or classrooms (Marsh, 1984, 1987; Marsh & Parker, 1984). Because ASC has been associated with many educational benefits, such as school persistence and achievement (Guay et al., 2004; Marsh et al., 2018; Szmuski & Karwowski, 2019), it is important to examine which school or classroom characteristics can attenuate the BFLPE.

The goal of the present study was to examine whether the CASIS professional development (PD) program could attenuate the BFLPE in relation to elementary students' writing ASC. The CASIS-PD program is anchored in Self-Determination Theory (SDT; Ryan & Deci, 2017) and aims to encourage teachers to rely on a set of five pedagogical practices that should increase students' academic motivation and achievement. We selected the writing school subject because a recent meta-analysis indicates that the BFLPE is stronger when ASC encompasses a verbal dimension such as reading or writing (Fang et al., 2018). Writing ASC has also been a key focus of inquiry in previous BFLPE studies (e.g., Huguet et al., 2009). This study is likely to contribute to the existing knowledge because very few studies have ever been able to document contextual moderators of the BFLPE. Yet, identifying school-related moderators of the BFLPE could provide valuable insights into teaching practices that could attenuate this undesirable effect.

The BFLPE

The BFLPE is rooted in students' perceptions of their abilities relative to those of other students (i.e., external frame of reference; Marsh, et al., 2020). More precisely, although ASC is known to correlate positively with individual levels of achievement, the BFLPE describes the fact that ASC also shares negative correlations with average classroom or school levels of achievement. In other words, if two students have the very same level of individual achievement, the one attending a classroom with high-achieving peers will develop a lower ASC comparatively to the one attending a classroom with peers achieving at a lower degree. This pattern of relation between the average level of achievement of the classroom or the school and ASC (i.e., the BFLPE) has been supported by a large body of international research (e.g., Marsh & Hau, 2003; Seaton et al., 2009; Nagengast & Marsh, 2012) and is one of the most strongly established findings in educational psychology (Stäbler et al., 2017).

At the heart of the BFLPE lies an imbalance between two social comparison processes: deliberate and forced (Huguet et al., 2009). Although deliberate social comparisons can sometimes be adaptative, forced social comparisons are primarily detrimental and assumed to represent the main driver of the negative association between class or school-average levels of achievement and students' ASC encapsulated in the BFLPE. On the one hand, students can make strategic social comparisons with self-selected targets that have slightly better grades (i.e., deliberate comparisons), and these comparisons may have a beneficial effect on subsequent performance and ASC through a process of assimilation (Huguet et al., 2001; Marsh et al., 2010). On the other hand, feedback on students' performance is also provided by teachers in relation to the achievement levels of the whole class, meaning that students are forced to position themselves in relation to all of their (possibly stronger) classmates (i.e., forced comparisons). This creates a situation where the negative effects of forced social comparisons on ASC counterbalance the potential positive effects of deliberate social comparisons (Huguet et al., 2009). Consistent with this

core component of the BFLPE, the negative effects of school-average levels of achievement on ASC has been shown to disappear once the effects of class-average levels of achievement are considered (Marsh et al., 2014). Moreover, a recent study showed that class-average levels of achievement was a more relevant source of information for students than the average level of achievement of their close peers (Koivuhovi et al., 2020). Class-average levels of achievement thus emerge as a source of information that students rely on, although not volitionally, to form their ASC in a specific school subject (Huguet et al., 2009).

Moderators of the BFLPE

Given that teachers are responsible for providing feedback on students' performance, it might be possible for some pedagogical practices to moderate the BFLPE. This question about whether the BFLPE varies across diverse educational settings has represented a key inquiry for BFLPE research from its inception. One of the first systematic attempts to address the issue of BFLPE moderation by contextual factors was realized by Lüdtke et al. (2005), who examined the influence of teachers' frame of reference on students' ASC. They proposed that the negative effect of class-average levels of achievement on students' ASC would be smaller when teachers used an individual reference standard (i.e., focusing on improvement, effort, and learning) rather than a social reference one (i.e., focusing on comparison between students). Even though their results revealed that students whose teacher used an individual reference standard had higher levels of ASC, this practice did not moderate the negative effect of class-average levels of achievement on ASC (i.e., the BFLPE). Conversely, Roy et al. (2015) demonstrated that differentiated instruction strategies (providing individualized feedback, varying teaching to match each student's learning needs) seemed to act as a moderator of the BFLPE. Indeed, they showed that the negative effect of class-average levels of achievement on ASC in French was attenuated for low-achieving students whose teacher did use differentiated instruction strategies. Although these results were limited to a subset of students (i.e., low achieving), this study provided evidence regarding the fact that teaching practices focusing on students' individual levels of mastery could moderate the BFLPE.

Beyond pedagogical practices meant to decrease imposed social comparisons by placing the focus on students' individual experience, the quality of the teacher-student relationships has also been identified as a contextual moderator of the BFLPE. Indeed, Schwabe et al. (2019) demonstrated that students who held positive relationships with their teachers (i.e., feeling understood, connected and at ease with them) were less affected by the BFLPE than students who perceived their relationship with their teacher to be average or poor. Positive relationships with teachers thus acted as a protective factor against the negative effect of class-average levels of achievement on students' ASC. However, this study also revealed that positive relationships with classmates did not moderate the BFLPE, even though classmates represent the very source of social comparison at the heart of the BFLPE. This result emphasizes the important role that teachers can have in promoting an educational environment that might act as a buffer against the negative effects of class-average levels of achievement on students' ASC. Overall, Schwabe et al.'s (2019) and Roy et al.'s (2015) results position teachers' pedagogical practices as a promising buffer against the negative effects of the BFLPE on students' ASC. However, more studies are needed to further clarify this moderating role of teacher-level variables in a way that can be harnessed to support intervention (e.g., Marsh et al., 2021). The present study pursues research on contextual moderators of the BFLPE using SDT as a framework.

Need-Supportive Practices: The CASIS Professional Development Framework

SDT (Ryan & Deci, 2017) proposes that the satisfaction of students' basic psychological needs for autonomy (feelings of volition), competence (feelings of competence or mastery), and relatedness (feeling of psychological closeness) is a key ingredient of optimal functioning at school. Indeed, the satisfaction of these needs has often been reported to play a role in relation to a variety of educational outcomes (e.g., motivation, achievement, engagement, etc.; see Ryan & Deci, 2017). Importantly, SDT acknowledges the critical role of need supportive pedagogical practices as a core driver of need satisfaction among students (Ryan & Deci, 2020). To integrate foundational work realized within and outside of SDT over the past decades (e.g., Guthrie et al., 2000; Reeve, 2002; Ryan & Deci, 2009; Skinner & Belmont, 1993), Guay et al. (2016; 2020) recently proposed five pedagogical practices that should contribute to creating learning conditions that would nurture and support students' basic psychological needs: Autonomy support, structure, involvement, significant activity, and cooperation.

When teachers use autonomy support practices, they consider the students' point of view, they give a rationale for requests, acknowledge students' feelings and perceptions, and provide them with

information and choices while minimizing pressure and control (e.g., performance-based rewards, competitive climate; Ryan & Deci, 2009). With structure, teachers set clear expectations, optimal challenges, and effective feedback for their students (Reeve, 2002). With involvement, teachers demonstrate a marked interest in students' lives, care about students' learning, and set realistic and positive goals (Skinner & Belmont, 1993). Teachers who are autonomy-supportive, use appropriate structure, and are involved with students may create a classroom climate where students' psychological needs are fulfilled. In addition to these three pedagogical practices, Guay et al. (2020) focuses on two others – significant activities and collaboration (Guthrie et al., 2000). Significant activities (or project-based learning) refer to meaningful writing activities which have real consequences for a child's life (Chen & Yan, 2019; Duke et al., 2006; Gambrell et al., 2011; Hiebert, 1994) and capture attention, raise questions, and promote active learning (Belet Boyaci & Güner, 2018; Boscolo & Gelati, 2013). With collaboration, students share their knowledge and ideas with their peers. Collaboration is associated with less competition among students (Guthrie et al., 2000) and with more positive attitudes toward writing (Li et al., 2014). Significant activities and collaboration are both central components of a classroom climate supportive of students' psychological needs. The former enables participation in activities students find interesting and enjoyable, which foster the satisfaction of their need for autonomy (Stroet et al., 2013). The latter promotes cohesiveness and positive interactions between students, thus positively impacting their need for relatedness (Sergis et al., 2018).

More importantly, to help teachers master these need supportive practices, Guay et al. (2020) proposed the CASIS (an acronym formed by the five practices: Collaboration, Autonomy-support, Structure, Involvement, and Significant activities) PD program. Bridging the gap between theory and practice, the general goal of this PD program is to encourage teachers to use evidence-based pedagogical practices in order to foster their students' academic motivation and achievement. With this in mind, Guay et al. (2020) tested the effectiveness of the CASIS-PD program in fostering students' autonomous forms of writing motivation. Although they found that this program promoted the use of need supportive practices by teachers, the effects on students' motivational resources were limited and moderated by their socioeconomic background. In the present study, we extended the work of Guay et al. (2020) by examining the moderating effect of the CASIS-PD program on the BFLPE, a more subtle phenomenon occurring across two levels of analysis (i.e., the student and the school environment). Because of the multilevel nature of the BFLPE, we can expect the role played by the CASIS-PD program, implemented at the teacher level, to be clearer than via a strict focus on student level outcomes. More precisely, we expect the CASIS-PD program to moderate the BFLPE by promoting an optimal learning context characterized by (1) a reduction in forced social comparisons involving the achievement level of the whole class (Huguet et al., 2009; Roy et al., 2015), and of (2) more positive teacher-student relationships (Schwabe et al., 2019). Indeed, by placing the focus on students' individual learning experience, interests, and intrapersonal development (Ryan & Deci, 2020; Chen & Yan, 2019), the autonomy support, structure, and significant activities practices are likely to substantially decrease forced social comparisons, which are a core component of the BFLPE (Huguet et al., 2009). Furthermore, the collaboration practice encourages students to share goals with some of their classmates, thus promoting group-based standards of collective achievement (Johnson & Johnson, 1999). Finally, by allowing students to feel relationally connected and emotionally supported by their teachers, which might be of particular importance when facing school-related learning challenges (Ryan & Deci, 2017), the involvement practice is likely to increase their likelihood of experiencing positive teacher-student relationships.

The Present Study

The goal of this study was to assess the potential moderating role of the CASIS-PD program on the BFLPE observed among a sample of elementary school children. The CASIS-PD program encompasses four units[1] taught to groups of 8 to 12 teachers during regular school hours, but outside of their teaching periods (i.e., pedagogical day). Each unit is delivered by an experienced elementary school teacher who has been trained by the research team. This quasi-experimental study comprises pretest (i.e., October)

[1] In the first unit, teachers receive theoretical explanations regarding students' motivation types and their relevance to achievement. In the second and third units, the five pedagogical practices of the CASIS-PD program are described to teachers with the help of written case studies and illustrated examples (i.e., videos). During this phase of the PD program, teachers are asked to record and observe their own daily practices to support their learning process. The fourth unit focuses on teachers' learning consolidation and mastery of each practice.

and posttest (i.e., June) assessments of children writing ASC and achievement (see Figure 1 for an illustration of the study design). Although the teachers were followed for two consecutive school years, the sample of children exposed to these teachers differed between the first and second year of the study (see Year 1 and Year 2 in Figure 1). Indeed, teachers teach to new cohorts of children each school year, allowing us to consider four groups of children in this study: (1) Year 1 children of the control group (CG) whose teachers had not yet participated in the CASIS-PD program, (2) Year 1 children of the experimental group (EG) whose teachers had participated in the CASIS-PD program, (3) Year 2 children of the CG whose teachers participated in the CASIS-PD program in a delayed manner, and (4) Year 2 children of the EG whose teachers had participated in the CASIS-PD program during Year 1. Based on this design, we posited that a BFLPE would be statistically significant only for Year 1 students of the control group ($CG_{Year\ 1}$) at posttest (see T2 in Figure 1). Indeed, Year 1 children from the CG were the only students who were not exposed to teachers who had followed the CASIS-PD program. For the three other groups of students, we did not expect any significant BFLPE. To test these hypotheses more stringently, we controlled for writing ASC at pretest for each group of students (see T1 and T3 in Figure 1). To our knowledge, this is the first quasi-experimental study to investigate potential moderators of the BFLPE. Compared to previous studies of contextual moderators of the BFLPE (Roy et al., 2015; Schwabe et al., 2019), the design of the present study is more stringent as it controls for classroom and individual characteristics that could explain the findings. This study is also more ecologically valid than laboratory experiments as it takes place within real school settings.

Method

Participants and Procedure

The data used in this study come from a quasi-experimental longitudinal project, which was approved by the Research Ethics Committee of the original authors' institution (see Guay et al., 2020, Study 2). The original project included elementary students enrolled in grades 1 to 6. However, because the BFLPE is usually stronger among older elementary school students (Televantou et al., 2021), we had to select students who were old enough to rely on social comparison processes to build their ability beliefs (DavisKean et al., 2009). For this reason, we rely on students enrolled in grades 4 to 6 (i.e., aged 9 to 12 years in Quebec's educational system).

During spring 2013 and 2014, elementary school teachers were recruited from three different school boards located in the Quebec City area (Quebec, Canada), deserving a predominantly white French-speaking population. Initially, 66 teachers from 50 schools agreed to participate. These 50 schools were randomly assigned to the experimental group (EG) or control group (CG). After the random assignment of schools, 19 teachers from the CG dropped out of the study. Of the 47 remaining teachers, 34 were teaching in the fourth, fifth, or sixth grade. Therefore, the present study focuses on 20 EG teachers from 15 different schools and 14 CG teachers from 9 schools. The demographic characteristics of these teachers are presented in Table 1. Teachers from both groups were followed for two consecutive school years. Teachers from the EG received the CASIS-PD program between T1 and T2 of the first year, while teachers from the CG received the CASIS-PD program between T3 and T4 of the second year (see Figure 1). CG teachers did not receive any alternative PD program during the first year of the study. Treatment fidelity was assessed by Guay et al. (2020), with the help of two independent judges who were blind to teachers' group allocation. All teachers, including CG teachers, were individually videotaped twice at each time-point during a writing lesson. Their level of application of the five CASIS pedagogical practices was evaluated using a Q-sort measure referring to 23 specific teaching behaviors. The results showed that teachers who received the CASIS-PD program applied these practices more often those who did not receive this PD program (for more detail on the treatment fidelity assessment procedure and results, see Guay et al., 2020).

As stated previously, students were followed during a single year only. Therefore, Year 1 students were different than Year 2 students within both the EG and CG. For the first year of the study, the EG and CG respectively included 394 and 190 students. For the second year of the study, the EG and CG respectively included 344 and 189 students. The students were therefore divided into four distinct groups: EG_{Year1}, CG_{Year1}, EG_{Year2}, and CG_{Year2} (see Figure 1). Of these groups, only the students of the CG_{Year1} were not exposed to teachers who had received the CASIS-PD program. Table 2 presents the students' characteristics. Teacher retention and student participation rates over the course of the study are reported in Table 3.

Measures

Writing Achievement

Teachers completed an in-house measure to assess students' writing achievement at pretest. Teachers were asked to evaluate each student's writing abilities against six thoroughly defined criteria rated on a scale ranging from 0 to 100%. These criteria were writing ideas, text organization, calligraphy, spelling, writing strategies, and text form. Scale score reliability (α) for this measure was .92 at T1 and .93 at T3.

Writing ASC

We used three positively worded items from the Self-Description Questionnaire – I (Marsh, 1990) to assess students' writing ASC. These items were: "I have always done well in writing", "Writing is easy for me", and "I learn quickly in writing". Students answered these items on a 5-point Likert-type scale (1 = *always no* to 5 = *always yes*) and completed them at pretest and posttest. The scale score reliability, convergent, and divergent validity of these three items were supported in a previous study (Guay et al., 2010). In the present study, scale scores reliability estimates (α) were .76 (T1), .83 (T2), .77 (T3), and .81 (T4).

Analyses

All analyses were realized using the robust maximum likelihood estimator (MLR) available in Mplus 8.3 (Muthén & Muthén, 2017), and full information maximum likelihood (FIML) estimation to handle the limited number of missing data present in our measures (T1: .2% to 7.2%, M = 3.28%; T2: 5.6% to 5.8%, M = 5.67%; T3: .6% to 16.9%, M = 8.2%; T4: 8.3% to 8.4%, M = 8.33%). Measures of writing ASC and writing achievement were standardized (M = 0, SD = 1) within each time point to facilitate interpretation and to avoid estimation problems linked to the widely discrepant rating scales of these measures (ASC 1-5; achievement: 0-100%) (e.g., Marsh et al., 2012; Morin et al., 2014). Across all models, model fit was assessed through an examination of the comparative fit index (CFI), the Tucker-Lewis Index (TLI), and the root mean square error of approximation (RMSEA), with CFI and TLI values ≥ .95 or .90, and RMSEA values ≤ .06 or .08 respectively used to indicate excellent and acceptable levels of fit to the data (Hu & Bentler, 1999; Marsh et al., 2005).

Measurement Model

Two baseline measurement models were estimated to ensure that our a priori factor structure fit the data correctly before adding any additional constraint or predictive path. The first of those models was a single level multi-group (across the four groups of students: EG_{Year1}, CG_{Year1}, EG_{Year2}, and CG_{Year2}) CFA model encompassing three factors (i.e., ASC at the beginning and end of the school year, and achievement at the beginning of the school year) freely estimated in the four groups. The second of those models was a doubly latent multilevel (with students at the first level, or L1, and classroom at the second level, or L2) multi-group (across the same four) CFA model encompassing the same factors freely estimated across L1 and L2 and the four groups (see Figure 2).

The adoption of a doubly latent approach means that (Marsh et al., 2009, 2012; Morin et al., 2014): (a) these latent factors were directly estimated from ratings provided by the students on each respective item, providing a way to control for measurement errors at L1 and L2, and that (b) these individual ratings were disaggregated into a L1 and L2 component via a latent aggregation procedure, providing a correction for sampling error in the estimation of the L2 reality. We used three item-parcels (Little et al., 2013) in the assessment of writing achievement to limit the number of free parameters. These parcels resulted from the combination of items referring to similar types of writing skills: Text organization and text form (Parcel 1), spelling and writing strategies (Parcel 2), calligraphy and writing ideas (Parcel 3). All factors were allowed to correlate with one another, and *a priori* correlated uniquenesses were included at L1 among matching indicators of the ASC factor across time points to avoid inflated stability estimates (Marsh, 2012).

After confirming the fit of the baseline models, we proceeded to tests of measurement invariance across time (for the ASC measures) and groups (for all constructs), to verify that the measurement structure of all constructs could be considered to be equivalent across groups and time (e.g., Millsap, 2011). These tests were conducted in the following sequence: (1) configural invariance (assuming the same factor structure over groups and time); (2) weak invariance (equivalence of the factor loadings over groups and time); (3) strong invariance (equivalence of the items' intercept over groups and time); (4) strict invariance (equivalence of the items' uniquenesses over groups and time). These tests were first conducted within a single-level measurement model, to verify that measurement structure of students' ratings of their own individual reality was equivalent across groups and time. For these single-level

models, a fifth step was used to verify the invariance of the a priori correlated uniquenesses across groups. Then, the most invariant of these models was converted into a doubly latent multi-group multilevel models, already assuming measurement invariance at L1[2], to test the measurement invariance of the L2 measurement structure according to the same four steps listed above.

Finally, to ensure that the constructs retained the same meaning across levels, we proceeded to a final test of measurement *isomorphism* (i.e., equality of the factor loadings across L1 and L2) starting from the most invariant model retained previously (e.g., Lüdtke et al., 2011; Morin et al., 2014). In addition to helping to stabilize model estimation by introducing more parsimony (Lüdtke et al., 2011), measurement *isomorphism* is an important prerequisite to our ability to consider the L1 constructs as a random variable with L2 variation (Morin et al., 2014, 2021), which is itself necessary to the calculation of the BFLPE (Marsh et al., 2012; Morin et al., 2014). Throughout these tests, increases in CFI and TLI $\geq$ .010 or decreases in RMSEA $\geq$ .015 between a more restricted model and the previous one in the sequence suggest that the less restricted model should be retained (Chen, 2007; Marsh et al., 2005).

Predictive Models

Finally, the retained doubly latent multilevel multi-group CFA model was converted to a doubly latent multilevel multi-group structural equation model (SEM) to assess whether the BFLPE would be present, and whether it would differ across (i.e., be moderated by) the four groups of students. For each group, we started by allowing writing ASC at posttest (T2 and T4) to be regressed on (i.e., predicted by) writing achievement and writing ASC at pretest (T1 and T3). In a first model (Model 1), these regressions were freely estimated across all groups of students. Then, a second model (Model 2) was estimated while constraining these predictions to be equal across all four groups of students. Finally, a last model (Model 3) was estimated, corresponding to our hypothesis, by constraining these paths to be equal across all groups of students whose teachers had participated in the CASIS-PD program (i.e., EG_{Year1}, EG_{Year2}, and CG_{Year2}), and allowed to differ between these groups and the remaining group of students whose teachers had not yet been exposed to the CASIS-PD program (i.e., CG_{Year1}).

Results from the retained model were used to calculate the BFLPE. Indeed, doubly latent models rely on an automatic group mean centering procedure, which is not adequate for the estimation of contextual effects (i.e., the L2 counterpart of a L1 effect relying on variables that reflect meaningful individual characteristics at L1, which is the case for the BFLPE; Marsh et al., 2012; Morin et al., 2014). In order to obtain proper estimates of contextual effects (i.e., including the BFLPE), it is necessary to subtract the L1 effect of writing achievement on writing ASC from its L2 counterpart (which corresponds to a grand-mean centered solution; Morin et al., 2014, 2021). Finally, all effects (including the properly calculated BFLPE) were properly standardized and converted to effects size indicators using formulas provided by Marsh et al. (2012) and Morin et al. (2014). We note that, based on these formulas, standardized coefficients and effects size indicators tend to be fairly similar to one another, especially for naturally individual level constructs. These effect size indicators can be interpreted following Cohen's guidelines (Cohen, 1988). Specifically, effect size values equal or greater than .10, .30, and .50 reflect small, moderate, and large effect sizes.

Results

The goodness-of-fit statistics associated with the various models estimated in this study are reported in Table 4. These results first revealed that both baseline models were able to achieve a satisfactory level of model fit (CFI/TLI $\geq$ .95; RMSEA $\leq$ .06). Likewise, all tests of measurement invariance were supported by the data, as none of these steps resulted in a decrease of model fit relative to the previous model in the sequence that exceeded the suggested guidelines (ΔCFI/TLI $\leq$ .010; ΔRMSEA $\leq$ .015). Although the multilevel model of configural invariance resulted in a suboptimal level of fit to the data, especially when considering the fit indices incorporating a correction for parsimony (TLI = .892 and RMSEA = .104), this misfit seems to be related to a lack of parsimony occurring when the L2 factor loadings had to be freely estimated over groups and time. Supporting this interpretation, the next model of invariance resulted in a substantial increase in model fit, resulting in an excellent level of fit to the data. Altogether, these results support the complete equivalence of the measurement structure across groups, time, and levels.

[2] With the exception of the item intercepts. Indeed, in doubly latent multilevel models, each item is associated with a single intercept, despite the fact that separate factor loadings and uniquenesses are estimated across levels. This means that intercepts had to be freed again, before testing the invariance of the L2 factor structure.

The last of these models was then converted to our a priori predictive model. As shown in the bottom section of Table 4, the model in which the predictive paths were constrained to be equal across all groups (Model 2) resulted in a substantial decrease in model fit (ΔCFI= -.027; ΔTLI= -.024; ΔRMSEA= +.021) relative to the model in which these paths were freely estimated across all groups (Model 1). In contrast, constraining these paths to be equal across the three groups of students whose teacher had received the CASIS-PD program, while allowing it to differ in the last group of students whose teachers had not yet been exposed to that program, resulted in a level of fit that was roughly equivalent to that of Model 1 (ΔCFI= -.002; ΔTLI= -.001; ΔRMSEA= .001). This more parsimonious model (i.e., Model 3), matching our theoretical expectations, was thus retained for interpretation. The results from this model are reported in Table 5 and summarized in Figure 3. These results revealed a statistically significant small to moderate BFLPE in the CG_{Year1} (standardized estimate = -.223, effect size = -.230), but not in any of the other groups, namely EG_{Year1}, EG_{Year2}, and CG_{Year2} (standardized estimate = from -.033 to -.037, effect size = from -.033 to -.037). In other words, a BFLPE was only observed among students whose teachers did not participate in the CASIS-PD program. Overall, these results provided further support to our expectations that the CASIS-PD program would be able to protect students against the BFLPE.

Discussion

The CASIS-PD program as a moderator of the BFLPE

The goal of this study was to test whether the CASIS-PD program for teachers, seeking to help them master practices intended to support students' basic psychological needs and aligned with SDT (Ryan & Deci, 2017), could help to protect elementary school students against the BFLPE. Contrasting with previous studies, the present study relied on a quasi-experiment in which the moderating variable was manipulated. Thus, the moderator (i.e., exposure to the CASIS-PD program) was less likely to be confounded with teachers or students' characteristics, thereby contributing to the reduction of potential bias in the estimation of the critical path coefficients. Moreover, initial levels of writing ASC were controlled for in the prediction of subsequent writing ASC, which made it possible to ensure that the observed effects of classroom levels of achievement were not a simple artifact of the effects of students' previous ASC on their levels of achievement. In addition, the doubly latent multilevel analyses applied in this study made it possible to obtain an estimate of the BFLPE that was controlled for measurement errors and sampling errors.

Our results provide support for the moderating role of the CASIS-PD program, revealing a BFLPE on writing ASC in the CG_{Year1}, but not in the other three groups of students (i.e., a BFLPE that was not statistically significant, and close to 0, in the EG_{Year1}, EG_{Year2}, and CG_{Year2}). Although the BFLPE obtained among students whose teachers were not exposed to the CASIS-PD program was small to moderate (Cohen, 1988), it was as strong as the BFLPE typically reported in the literature in relation to the verbal domain (Fang et al., 2018). Thus, our results suggest that training teachers in the use of the five pedagogical practices proposed by the CASIS-PD program, namely autonomy support, structure, involvement, collaboration, and significant activities, might be sufficient to counteract the BFLPE, or at least to reduce its magnitude. In addition to being aligned with past results showing that contextual variables can moderate the BFLPE (Roy et al., 2015; Schwabe et al., 2019), our results are consistent with SDT. Indeed, students who feel autonomous, competent, and connected with their teacher tend to endorse their learning activities, to believe that they have the potential to succeed, and to feel a strong sense of personal worth. Furthermore, pairing need-supportive practices with significant activities and collaboration should also help to promote students' learning proactivity and positive peer relationships (Guay et al., 2020). On this basis, our results suggest that exposure to an improved need-supportive teaching context, as a result of teachers' exposure to the CASIS-PD program, seem to help students focus on their own achievement and improvement to shape their ASC according to their own standards, rather than to rely extensively on social comparisons involving the achievement of the whole class. Such a need-supportive context might also protect children from the negative effect of class-average levels of achievement on their writing ASC when such comparisons occur.

Our findings have important practical implications. First, the CASIS-PD program seems to represent a promising avenue to attenuate the negative effects of the BFLPE, perhaps not only in written French, but also in other school subjects. Moreover, these five practices should be helpful for both low- and high-achieving students by reducing the negative effect of class-average levels of achievement on ASC no matter the average level of class achievement. Second, there is a global trend in education to

create special programs for more talented students (Marsh et al., 2014), which is also true in Quebec's educational system. The aim of these selective programs is to offer more challenging material and education to students to help them develop their full potential. The BFLPE indicates that the risks associated with such programs is that exposing students to high-achieving classrooms is likely to undermine their ASC, and in turn to negatively affect the many outcomes known to be strongly associated with ASC, such as school perseverance, academic achievement, and educational attainment (e.g., Guay et al., 2004; Marsh, 2007). For these reasons, we believe that the implementation of the CASIS-PD program should be an important part of these selective programs, allowing them to counteract this negative effect. Interestingly, no BFLPE was observed among EG_{Year2}, students even though their teachers participated in the CASIS-PD program in Year 1 (i.e., several months before). This therefore suggests that this PD program has lasting effects on the pedagogical practices adopted by teachers, an observation that should further encourage its implementation within selective programs, and within schools in general.

Achievement Measures Used to Evaluate the BFLPE

Previous research proposed that classmates represent a better frame of reference than students from the entire school or from small groups of peers (Huguet et al., 2009; Koivuhovi et al., 2020). In line with these past studies, we also observed (for students whose teachers had not been exposed to the CASIS-PD program) that classmates' writing achievement was negatively associated with students' writing ASC even while controlling for their initial levels of writing ASC. However, contrasting with past BFLPE research, we did not use standardized test scores or report cards to measure achievement, but rather asked teachers to grade their students. Marsh et al. (2014) argued that report cards are suboptimal to test the BFLPE because teachers tend to grade on a curve, giving higher marks to the most skillful children and lower marks to the least skillful ones in their classes. As a result, the mean level of achievement observed in a class tends to be very similar across classrooms, in addition to resulting from the application of possibly distinct metrics. In contrast, standardized measures of achievement result in a more accurate assessment of children's levels of achievement in relation to developmentally appropriate standards. However, it is also true that class marks represent a more proximal source of feedback for students, and thus represent a stronger predictor of ASC than standardized test scores (e.g., Marsh, 2007; Marsh et al., 2014). Yet, results generally show that this contribution of class marks to ASC is independent from the BFLPE, and might even underestimate its magnitude (Marsh et al., 2014).

In this study, rather than relying on class marks or on standardized-test scores, we asked our sample of teachers to use a grid to rate the achievement level of their students on six writing competencies in relation to specific standards of success. This rating process was based on the writing activities carried out within each classroom. Consequently, the performance criteria used by the teachers were the same across classrooms, but the writing activities could differ from one classroom to another. Our measure was thus standardized in terms of its form (i.e., writing skills), but not in terms of its content (i.e., writing tasks). We believe, however, that the writing skills that we assessed encompass all writing tasks, meaning that they are transferable from one writing task to another (e.g., if a student has a specific writing skill in a specific writing task, he or she is more likely to have this skill in another writing task). Although not as accurate as standardized test scores, this criterion-referenced grading system seemed to represent a more appropriate way to test for the BFLPE than class marks given that it forces teachers to rate students in relation to specific (and objective) standards rather than in comparison with one another (Aviles, 2001; Martuza, 1977). Because it is not always possible to have access to a standardized test to assess the BFLPE, the measure used in this study could be a valid alternative to test scores. However, rigorous testing of the validity of this measure in comparison with class marks and standardized achievement tests should be the object of future research. Likewise, it would also be important to replicate the present findings in relation to other subject areas in order to more specifically assess the transferability of the CASIS-PD program.

Strengths, Limitations, and Future Directions

Among the numerous strengths of this study is the quasi-experimental pretest-posttest design and random assignments of schools to the study conditions. Moreover, the treatment fidelity was assessed by external judges (see Guay et al., 2020), and we relied on a doubly latent approach to test our hypotheses while controlling for measurement and sampling errors. Together, these strengths lend confidence to our results. However, this study also has limitations that should be highlighted. First, it is not possible to know whether specific components of the CASIS-PD program, or specific practices covered in this program, were more important than others in moderating the BFLPE. Future studies could address this

issue by systematically assessing teachers' use of the five pedagogical practices included in CASIS to more directly link them with the BFLPE. This could be done by frequently recording teachers during lessons to have a more thorough portrait of classroom functioning. Second, the mechanisms by which the CASIS-PD program moderated the BFLPE are not known with certainty. Indeed, we did not assess students' reliance on social comparison processes, nor teachers' tendency to compare students' performance to that of the whole class. Similarly, we did not assess other variables assumed to play a role in the observed moderation effect (e.g., basic psychological needs). Thus, we can only rely on hypothetical interpretations to understand our results, and postulate that the moderation of the BFLPE observed in this study was explained by the fact that the CASIS-PD program promoted positive teacher-student relationships as well as an optimal learning context which helped students to focus more on themselves and their own experiences with French writing than on the performance level of the whole class. Future studies should also consider additional mechanisms (e.g., students' need satisfaction, teachers' reliance on forced social comparisons) that might contribute to further explain the moderation of the BFLPE by the CASIS-PD program. Third, our results are limited to the French writing domain and to the Quebec City area located in the Canadian province of Quebec (i.e., a predominantly white French-speaking area), suggesting the need to test the extent to which the present results would generalize to other academic domains, linguistic and ethnic groups, and cultures. Lastly, the BFLPE was not estimated using standardized test scores, which are the gold standard in this research field (Marsh et al., 2007). Although the criterion-referenced grading system used in this study might be a viable alternative when standardized test scores are not available (as noted above), our results need to be replicated using recommended measures of achievement.

Conclusion

By being negatively related to students' ASC, the BFLPE can be detrimental for educational outcomes such as achievement and attainment. Finding contextual moderators of the BFLPE that can contribute to decreasing the salience of social comparison processes for students' ASC thus becomes an important endeavor. In this quasi-experimental study, we provide empirical support for the ability of the CASIS-PD program to offset the BFLPE in relation to the French writing ASC of late elementary school children. It appears that the application of autonomy support, structure, involvement, significant activities, and collaboration by teachers is likely to foster an optimal learning context that encourages students to use their own improvements and accomplishments as benchmarks to shape their ASC.

References

Aviles, C.B. (2001). Grading with norm-referenced or criterion-referenced measurements: To curve or not to curve, that is the question. *Social Work Education*, *20*, 603–608.

Belet Boyaci, S.D., & Güner, M. (2018). The impact of authentic material use on development of the reading comprehension, writing skills and motivation in language course. *International Journal of Instruction*, *11*, 351–368.

Boscolo, P., & Gelati, C. (2013). Best practices in promoting motivation for writing. In S. Graham, C.A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (2nd Ed., pp. 284–308). Guilford.

Buunk, B.P., Kuyper, H., & van Der Zee, Y.G. (2005). Affective response to social comparison in the classroom. *Basic and Applied Social Psychology*, *27*, 229–237.

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

Chen, C.H., & Yang, Y.C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review*, *26*, 71-81.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Davis-Kean, P.E., Jager, J., & Andrew Collins, W. (2009). The self in action: An emerging link between self-beliefs and behaviors in middle childhood. *Child Development Perspectives*, *3*, 184-188.

Dijkstra, P., Kuyper, H., Van der Werf, G., Buunk, A.P., & van der Zee, Y.G. (2008). Social comparison in the classroom: A review. *Review of educational research*, *78*, 828-879.

Duke, N.K., Purcell-Gates, V., Hall, L.A., & Tower, C. (2006). Authentic literacy activities for developing comprehension and writing. *The Reading Teacher*, *60*, 344–355.

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: a meta-analysis. *Frontiers in Psychology*, *9*, 1569.

Gambrell, L.B., Hughes, E.M., Calvert, L., Malloy, J.A., & Igo, B. (2011). Authentic reading, writing, and

discussion: An exploratory study of a pen pal project. *The Elementary School Journal*, *112*, 234–258.
Guay, F., Gilbert, W., Falardeau, É., Bradet, R., & Boulet, J. (2020). Fostering the use of pedagogical practices among teachers to support elementary students' motivation to write. *Contemporary Educational Psychology*, *63*, 101922.
Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity*, *3*, 53–68.
Guay, F., Valois, P., Falardeau, E., & Lessard, V. (2016). Examining the effects of a professional development program on teachers' pedagogical practices and students' motivational resources and achievement in written French. *Learning and Individual Differences*, *45*, 291–298.
Guay, F., Ratelle, C.F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*, 644–653.
Guthrie, J.T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology*, *92*, 331–341.
Hiebert, E.H. (1994). Reading recovery in the United States: What difference does it make to an age cohort? *Educational Researcher*, *23*, 15–25.
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.
Huguet, P., Dumas, F., Marsh, H.W., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, *97*, 156–170.
Huguet, P., Dumas, F., Monteil, J.-M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, *31*, 557–578.
Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into Practice*, *38*, 67–73.
Koivuhovi, S., Marsh, H.W., Dicke, T., Sahdra, B., Guo, J., Parker, P.D., & Vainikainen, M.P. (2020). Academic self-concept formation and peer-group contagion: Development of the big-fish-little-pond effect in primary-school classrooms and peer groups. *Journal of Educational Psychology*. Advance online publication.
Li, X., Chu, S.K., & Ki, W.W. (2014). The effects of a wiki-based collaborative process writing pedagogy on writing ability and attitudes among upper primary school students in Mainland China. *Computers & Education*, *77*, 151–169.
Little, T.D., Rhemtulla, M., Gibson, K., Schoemann, A.M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*, 285-300.
Lüdtke, O., Köller, O., Marsh, H.W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology*, *30*, 263–285.
Lüdtke, O., Marsh, H.W., Robitzsch, A., & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*, 444-467.
Marsh, H.W. (1984). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal*, *21*, 799–806.
Marsh, H.W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*, 129–149.
Marsh, H.W. (1987). The big-fish–little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280–295.
Marsh, H.W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*, 623–636.
Marsh, H.W. (2007). *Self-Concept Theory, Measurement and Research into Practice*. British Psychological Society.
Marsh, H.W. (2012). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (3rd Ed., pp. 774 –798). Wiley.
Marsh, H.W., Byrne, B.M., & Shavelson, R.J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*,

80, 366–380.

Marsh, H.W., & Craven, R.G. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Academic Press.

Marsh, H.W., & Hau, K.T. (2003). Big-Fish—Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*, 364–376.

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* (pp. 275-340). Erlbaum.

Marsh, H.W., Kuyper, H., Morin, A.J.S., Parker, P.D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50-66.

Marsh, H.W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764-802.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A.J.S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*, 106–124.

Marsh, H.W., & Parker, J.W. (1984). Determinants of students self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*, 213–231.

Marsh, H.W., Parker, P.D., Guo, J., Pekrun, R., Basarkod, G., & Kandler, C. (2020). Psychological comparison processes and self-concept in relation to five distinct frame–of–reference effects: Pan–human cross–cultural generalizability over 68 countries. *European Journal of Personality, 34*, 180-202.

Marsh, H.W., Pekrun, R., Murayama, K., Arens, A.K., Parker, P.D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*, 263-280.

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Régner, I., Buunk, A. P., Monteil, J.-M., & Gibbons, F. X. (2010). Phantom behavioral assimilation effects: Systematic biases in social comparison choice studies. *Journal of Personality*, *78*, 671-710.

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*, 631-669.

Marsh, H. W., Xu, K. M., Parker, P. D., Hau, K. T., Pekrun, R., Elliot, A., Guo, J., Dicke, T., & Basarkod, G. (2021). Moderation of the Big-Fish-Little-Pond Effect: Juxtaposition of Evolutionary (Darwinian-Economic) and Achievement Motivation Theory Predictions Based on a Delphi Approach. *Educational Psychology Review*, 1-26.

Martuza, V.R. (1977). *Applying norm-referenced and criterion-referenced measurement in education*. Allyn & Bacon.

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

Morin, A. J., Blais, A. R., & Chénard-Poirier, L. A. (2021). Doubly Latent Multilevel Procedures for Organizational Assessment and Prediction. *Journal of Business and Psychology*, 1-26.

Morin, A.J.S., Marsh, H.W., Nagengast, B., & Scalas, L.F. (2014). Doubly Latent Multilevel Analyses of Classroom Climate: An illustration. *Journal of Experimental Education*, *82*, 143-167.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Nagengast, B. & Marsh, H.W. (2012). Big fish in little ponds aspire more: mediation and cross-cultural generalizability of school-average ability effects on self0concept and career aspirations in science. *Journal of Educational Psychology*, *104*, 1033-1053.

Reeve, J. (2002). Self-determination theory applied to educational settings. In E.L. Deci, & R.M. Ryan (Eds.), *Handbook of self-determination research* (pp. 183–203). University of Rochester.

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110–116.

Ryan, R.M., & Deci, E.L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171–196). Routledge.

Ryan, R.M., & Deci, E.L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860.

Schwabe, F., Korthals, R., & Schils, T. (2019). Positive social relationships with peers and teachers as moderators of the Big-Fish-Little-Pond Effect. *Learning and Individual Differences*, *70*, 21-29.

Seaton, M., Marsh, H.W., & Craven, R.G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*, 403–419.

Sergis, S., Sampson, D. G., & Pelliccione, L. (2018). Investigating the impact of Flipped Classroom on students' learning experiences: A Self-Determination Theory approach. *Computers in Human Behavior*, *78*, 368-378.

Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407–441.

Skinner, E.A., & Belmont, M.J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581.

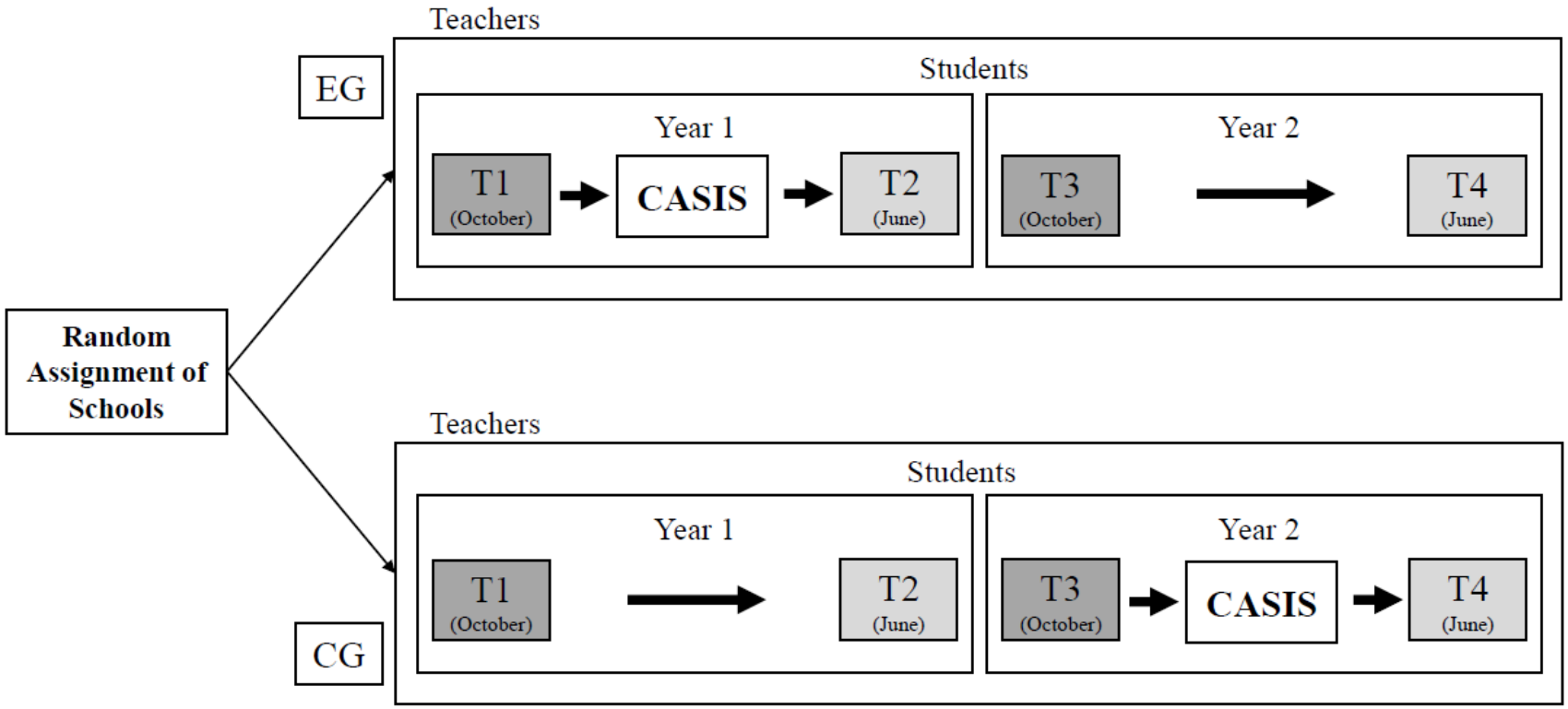
Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology*, *109*, 191–207.

Stroet, K., Opdenakker, M. C., & Minnaert, A. (2013). Effects of need supportive teaching on early adolescents' motivation and engagement: A review of the literature. *Educational Research Review*, *9*, 65-87.

Szumski, G., & Karwowski, M. (2019). Exploring the Pygmalion effect: The role of teacher expectations, academic self-concept, and class context in students' math achievement. *Contemporary Educational Psychology*, *59*, 101787.

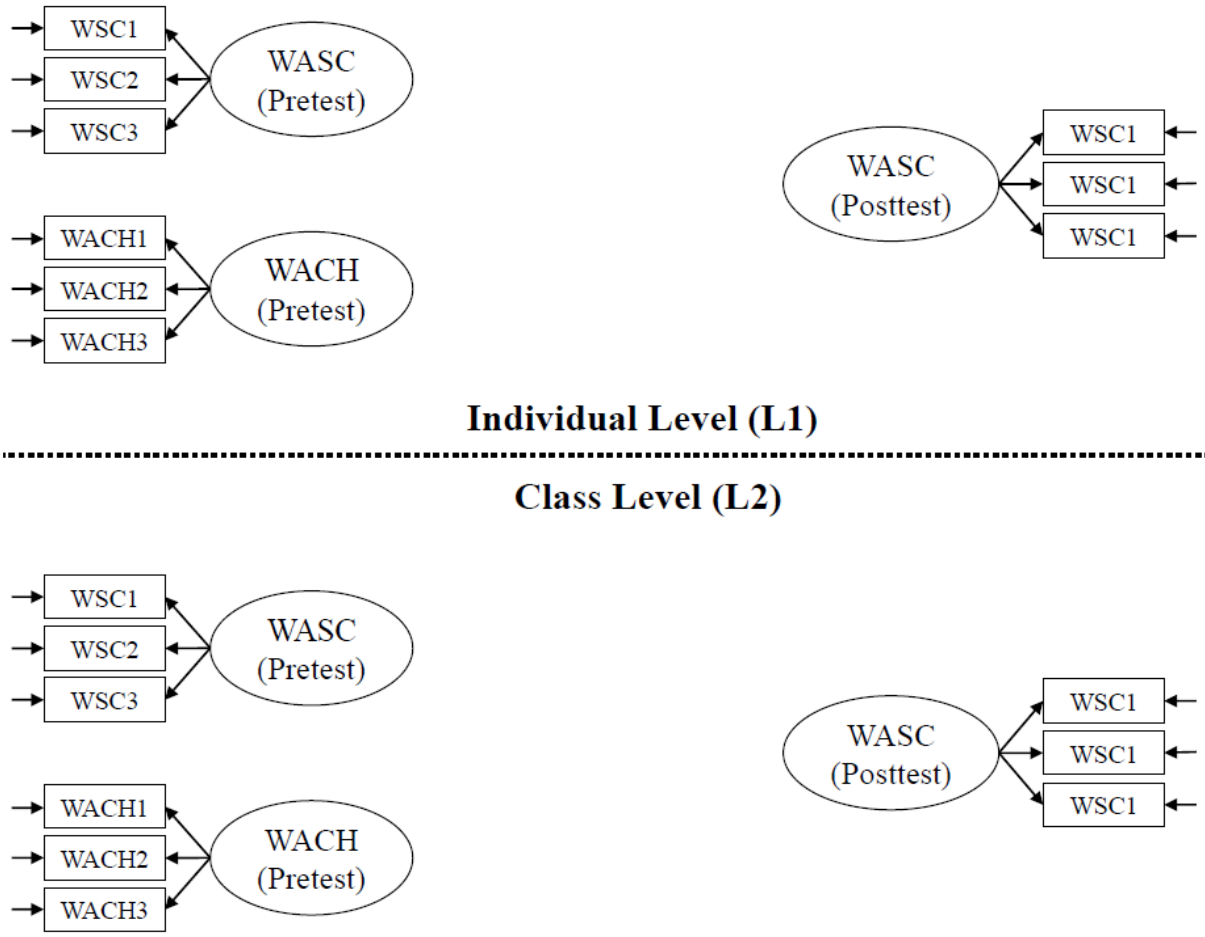
Televantou, I., Marsh, H. W., Dicke, T., & Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement: Evidence from English early primary schools. *Learning and Instruction*, *71*, 101399.

Figure 1
Illustration of the study design



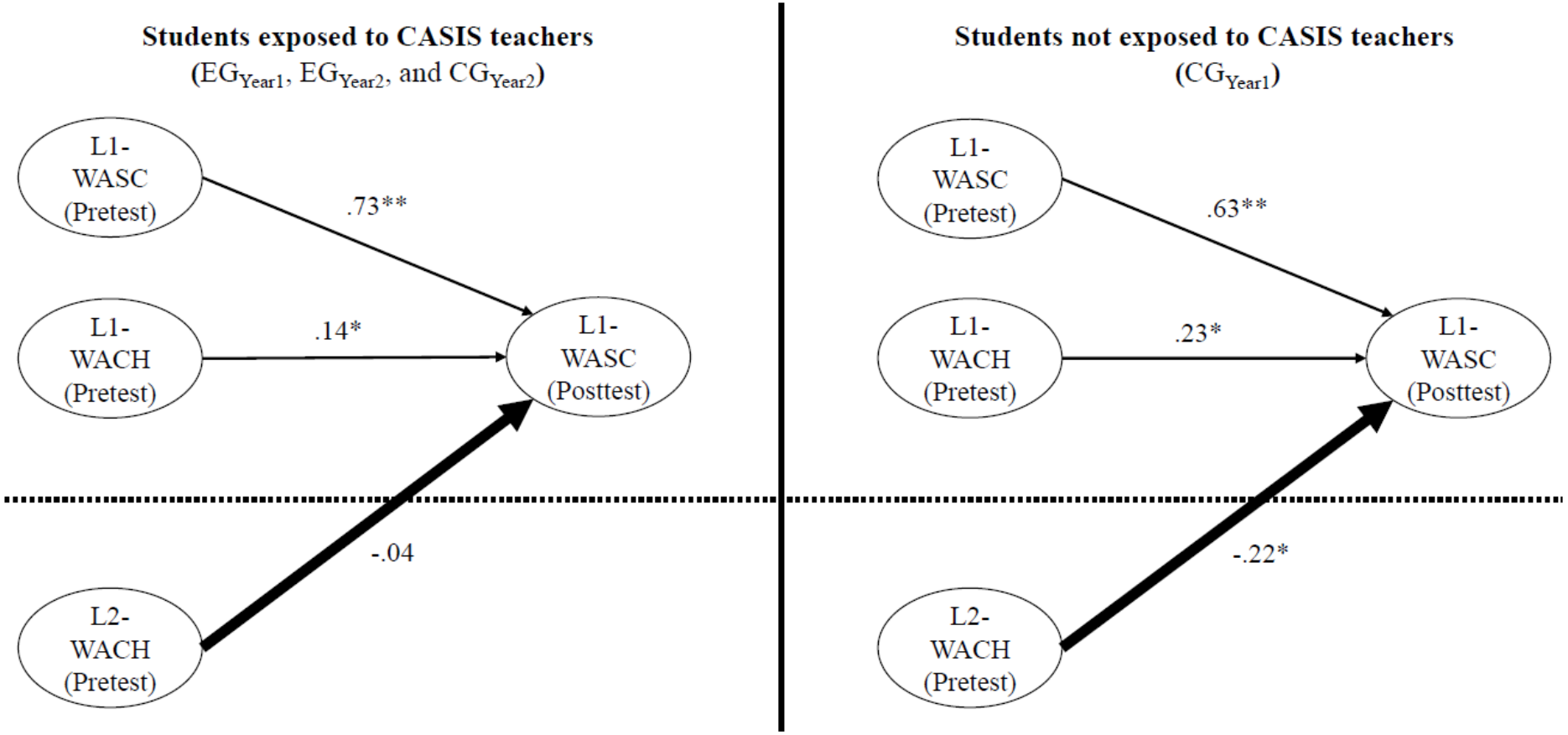
Note. EG = Experimental group; CG = Control group; T1= Pretest of Year 1; T2 = Posttest of Year 1; T3 = Pretest of Year 2; T4 = Posttest of Year 2. This study was composed of two groups of teachers and four groups of students. Students' grade level is not depicted in this figure.

Figure 2
Illustration of the baseline multilevel CFA model



Note. This model encompassed the four groups of students and did not include predictive paths. WASC = Writing academic self-concept; WACH = Writing achievement; Ovals represent latent factors; Rectangles represent observed indicators.

Figure 3
Illustration of the Big-Fish-Little-Pond-Effect



Note. This model is a simplification of Figure 2. The BFLPE is represented by the thickest arrow. Each number shown on the left side of the figure is the mean of the results from the EG$_{Year1}$, EG$_{Year2}$, and CG$_{Year2}$. Correlations between predictors are not shown for purpose of simplicity. L1 = Individual level; L2 = Classroom level; WASC = Writing academic self-concept; WACH = Writing achievement; Ovals represent latent factors.
* = $p < .05$; ** = $p < .01$.

Table 1

Teachers' Characteristics

	EG			CG		
Variables	*n*	%		*n*	%	
Gender						
Male	2	10.0		1	92.9	
Female	18	90.0		13	7.1	
Total	20	100		14	100	
Grades						
4	6	30.0		4	28.6	
5	10	50.0		5	35.7	
6	4	20.0		5	35.7	
Total	20	100		14	100	
Educational level						
Bachelor	15	75.0		9	64.3	
Master	5	25.0		5	35.7	
Total	20	100		14	100	
	n	Mean	*SD*	*n*	Mean	*SD*
Age	20	38	5.14	14	38.07	7.46

Note. EG = Experimental group; CG = Control group.

Table 2

Students' Characteristics

	Year 1				Year 2			
	EG		CG		EG		CG	
Variables	*n*	%	*n*	%	*n*	%	*n*	%
Gender								
Male	180	45.7	87	45.8	168	48.8	102	54.0
Female	214	54.3	103	54.2	176	51.2	87	46.0
Total	394	100	190	100	344	100	189	100
Grades								
4	129	32.7	58	30.5	59	17.2	52	27.5
5	151	38.3	66	34.7	211	61.3	93	49.2
6	114	28.9	66	34.7	74	21.5	44	23.3
Total	394	100	190	100	344	100	189	100
Mothers' education level								
High school incomplete	11	7.0	12	14.6	8	6.6	4	5.3
High school completed	16	10.3	14	17.1	14	11.5	12	16.0
College completed	69	44.2	32	39.0	47	38.5	31	41.3
Undergraduate studies	43	27.6	18	22.0	32	26.2	19	25.3
Master or Ph.D.	17	10.9	6	7.3	21	17.2	9	12.0
Missing	238	-	108	-	222	-	114	-
Total	394	100	190	100	344	100	189	100
Family income								
Less than C$40,000	27	15.3	21	24.1	16	11.3	16	19.5
C$40,000-69,999	51	28.8	26	29.9	32	22.5	24	29.3
C$70,000 or more	99	55.9	40	46.0	94	66.2	42	51.2
Missing	217	-	103	-	202	-	107	-
Total	394	100	190	100	344	100	189	100
Language spoken at home								
French	175	98.3	83	95.4	137	96.5	78	95.1
Other	3	1.7	4	4.6	5	3.5	4	4.9
Missing	216	-	103	-	202	-	107	-
Total	394	100	190	100	344	100	189	100
Lives with both biological parents								
Yes	136	76.4	61	70.1	112	78.9	56	68.3
No	42	23.6	26	29.9	30	21.1	26	31.7
Missing	216	-	103	-	202	-	107	-
Total	394	100	190	100	344	100	189	100

Note. EG = Experimental group; CG = Control group.

Table 3

Teachers Retention and Student Participation throughout the Study

	Year 1			Year 2		
Participants	T1	T2	RR (%)	T3	T4	RR (%)
Teachers - EG	20	20	100	18	18	100
Teachers - CG	14	14	100	13	11	84.61
Teachers - Total	34	34	100	31	29	93.55
Students - EG	394	382	96.95	344	342	99.42
Students - CG	190	195	102.63	189	175	92.59
Students - Total	584	577	98.80	533	517	97

Note. EG = Experimental group; CG = Control group; RR = Retention rate.

Table 4

Fit Indices for Each Model Estimated in the Present Study

Model	χ^2	df	CFI	TLI	RMSEA
Baseline Multi-Group CFA Models					
Single Level	142.966	78	.994	.988	.031
Multilevel	232.345	165	.993	.988	.035
Single Level Multi-Group CFA Models					
Configural invariance	142.966	78	.994	.988	.031
Weak invariance	166.535	102	.994	.991	.027
Strong invariance	154.008	120	.997	.996	.018
Strict invariance	171.335	153	.998	.998	.012
Correlated Uniquenesses Invariance	179.217	162	.998	.998	.011
Multilevel Multi-Group CFA Models					
Configural invariance	1043.37	228	.914	.892	.104
Weak invariance	511.72	252	.973	.969	.056
Strong invariance	501.83	270	.976	.974	.051
Strict invariance	550.54	297	.973	.974	.051
Measurement isomorphism	547.91	303	.974	.975	.049
Predictive Models					
Model 1	529.31	303	.976	.977	.047
Model 2	801.50	315	.949	.953	.068
Model 3	553.36	311	.974	.976	.048

Note. Although Maximum Likelihood (ML) chi-square and CFI should be monotonic with model complexity, their robust (MLR) counterpart can improve with added constraints as a result of differences in the scaling correction factor. χ^2 = 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.

Table 5
Estimation of the Big-Fish-Little-Pond-Effect

	Predictive paths to WASC			
Predictors	*B*	*SE B*	*β*	ES
EG students – Year 1				
Level 1				
WASC	.807**	.061	.730	.731
WACH	.120*	.011	.132	.132
Level 2				
WASC	.109	.320	.020	.020
WACH	.032	.109	.013	.013
BFLPE	-.087	.12	-.035	-.035
CG students – Year 1				
Level 1				
WASC	.698**	.135	.629	.647
WACH	.220*	.103	.231	.238
Level 2				
WASC	1.294**	.286	.268	.276
WACH	-.229	.131	-.114	-.117
BFLPE	-.450*	.17	-.223	-.230
EG students – Year 2				
Level 1				
WASC	.807**	.061	.754	.755
WACH	.120*	.047	.136	.136
Level 2				
WASC	.109	.320	.020	.020
WACH	.032	.109	.014	.014
BFLPE	-.087	.12	-.033	-.033
CG students – Year 2				
Level 1				
WASC	.807**	.061	.716	.733
WACH	.120*	.047	.136	.139
Level 2				
WASC	.109	.320	.024	.025
WACH	.032	.109	.012	.012
BFLPE	-.087	.12	-.037	-.037

Note. For Year 1 students, predictors were measured at T1 and the outcome at T2. For Year 2 students, predictors were measured at T3 and the outcome at T4; EG = Experimental group; CG = Control group; WASC = Writing academic self-concept; WACH = Writing achievement; BFLPE = Big-Fish-Little-Pond effect; *B* = Unstandardized coefficient*; SE B* = Standard error for the unstandardized coefficient; *β* = Standardized coefficient; ES = Effect size.
* = $p < .05$; ** = $p < .01$.