Learning and Individual Differences 27 (2013) 109–119

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Do teacher judgments of student intelligence predict life outcomes?

Antoine Fischbach [a,*], Tanja Gabriele Baudson [b], Franzis Preckel [b], Romain Martin [c], Martin Brunner [d]

[a] University of Luxembourg, EMACS Research Unit, 6, rue Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg
[b] University of Trier, Department of Psychology (FB I), D-54286 Trier, Germany
[c] University of Luxembourg, EMACS Research Unit, B.P. 2, L-7201 Walferdange, Luxembourg
[d] Free University of Berlin and Berlin-Brandenburg Institute for School Quality, Otto-von-Simson-Str. 15, D-14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form 31 May 2013
Accepted 12 July 2013

Keywords:
Teacher judgment
Self-fulfilling prophecy
Intelligence
IQ
Longitudinal

ABSTRACT

The present paper investigated how *teacher judgments* (TJs) of student intelligence at age 12 predicted adult intelligence and further key life outcomes (i.e., educational attainment, socioeconomic achievement, health, and subjective well-being) at age 52. Using path analytic techniques on a representative sample of $n = 731$ Luxembourg students, we found that TJs predicted key life outcomes (except subjective well-being) across a time span of 40 years. Findings remained stable even after controlling for childhood intelligence. When adjusting for GPA, the predictive power of TJs dropped considerably. This emphasizes that the predictive power of TJs hearkens back to the fact that TJs mainly reflect students' educational achievement rather than students' intelligence. Given that TJs had potential self-fulfilling effects on years of education, our data supported the hypothesis that self-fulfilling prophecies may have small long-term effects on the development of adult intelligence and some key life outcomes that are mediated via educational attainment.

1. Introduction

Teachers need to judge their students' cognitive abilities in order to adapt their classroom instruction to their students' cognitive prerequisites (e.g., Brunner, Anders, Hachfeld, & Krauss, 2013). Moreover, *teacher judgments* (TJs) of students' abilities may also provide information with regard to their evaluations of student performance when grading and making placement decisions. Thus, TJs may affect student learning. Furthermore, TJs likely shape students' educational and occupational careers and related life outcomes (Afarian & Kleiner, 2003; Roth, BeVier, Switzer, & Schippmann, 1996; Roth & Clarke, 1998; Schuler, Funke, & Baron-Boldt, 1990; see also Strenze, 2007). The major objective of the present prospective study was to examine these predictions. To this end, we investigated how TJs of student intelligence at age 12 predicted adult intelligence and key life outcomes (i.e., educational attainment, socioeconomic achievement, subjective well-being, and health) at age 52.

2. Previous research on TJs

Previous research on TJs has focused on two interrelated aspects: the accuracy of TJs of students' abilities and the short- and midterm consequences of teachers' over- and underestimations of these abilities (i.e., misjudgments).

2.1. TJ accuracy

With regard to accuracy, Südkamp, Kaiser, and Möller (2012) recently meta-analyzed 75 studies that examined the correlation between TJs of students' cognitive achievement and students' actual performance on standardized achievement tests. The mean correlation was $r = .63$, demonstrating that higher TJs correspond to better performance on achievement tests, or in other words, that teachers' judgment accuracy is fairly high in general. Therefore, if TJs are predictive of student outcomes, it is first and foremost because these assessments are quite accurate indeed (Jussim & Harber, 2005; see also Brophy, 1983; Smith et al., 1998). Importantly, to adequately analyze TJ accuracy, it is essential that students' actual performance is understood as a standardized measure of performance that is independent of teacher influence. For example, school grades are a commonly used index of academic achievement; nonetheless, they are a form of TJ themselves (Alvidrez & Weinstein, 1999).

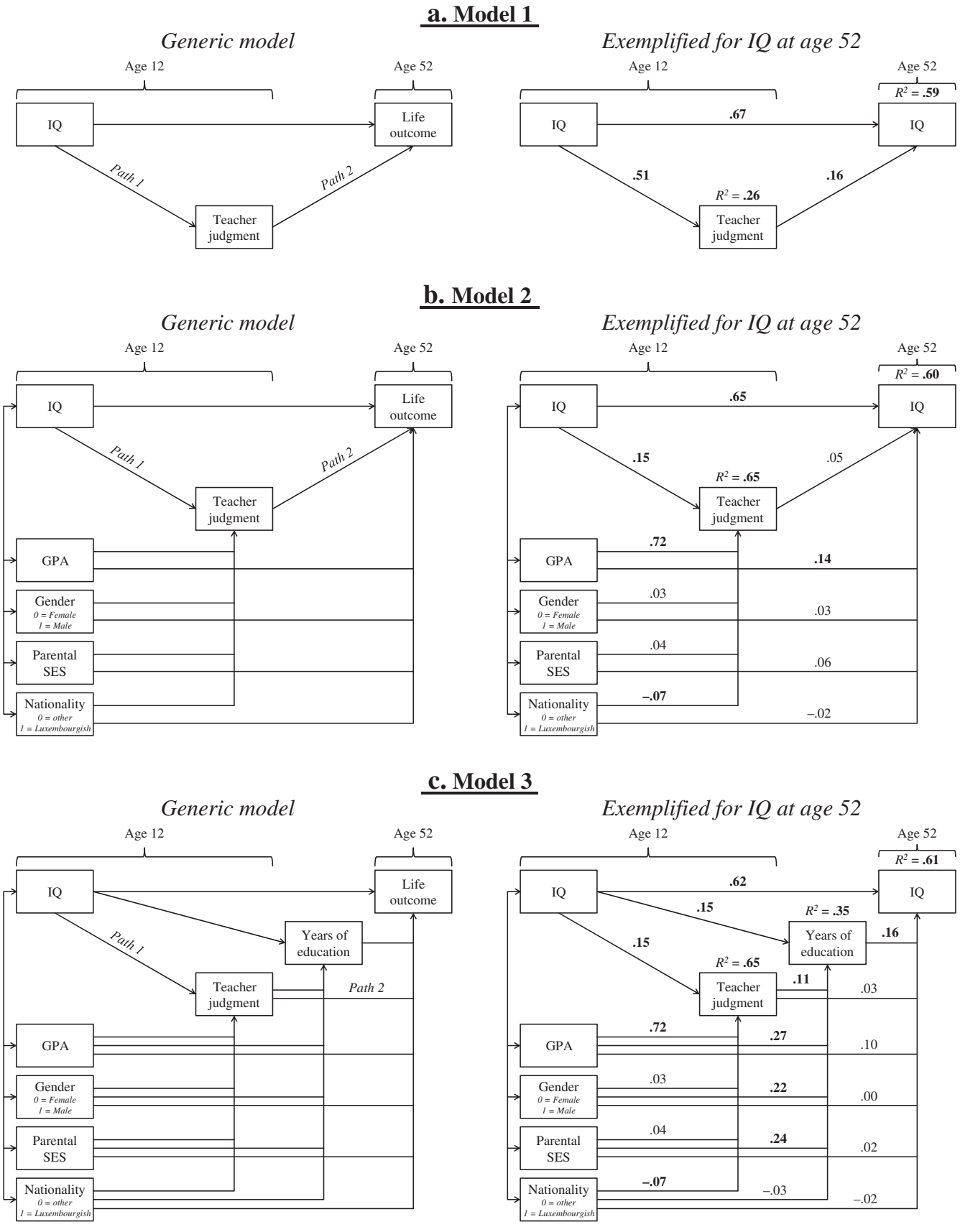
2.2. Consequences of misjudgment

The consequences of over- and underestimations of individual student characteristics have primarily been analyzed in the context of *self-fulfilling prophecies* (SFPs; Merton, 1948; see also Brophy, 1983; Rosenthal & Jacobson, 1968; Rosenthal & Rubin, 1978). Initial research on SFPs—the so-called *Pygmalion* study—focused on intelligence (Rosenthal & Jacobson, 1968). In this case, an SFP implies that teachers' overestimations of their students' intelligence lead to these students turning out brighter than would be expected given their actual levels of intelligence. On the other hand, when teachers underestimate their

* Corresponding author. Tel.: +352 466 644 9643; fax: +352 466 644 9556.
E-mail addresses: antoine.fischbach@uni.lu (A. Fischbach), baudson@uni-trier.de (T.G. Baudson), preckel@uni-trier.de (F. Preckel), romain.martin@uni.lu (R. Martin), martin.brunner@isq-bb.de (M. Brunner).

1041-6080/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.lindif.2013.07.004

students' intelligence, an SFP implies that the intellectual development of these students would be impaired. To investigate mediators of SFPs in classroom contexts, Harris and Rosenthal (1985) meta-analyzed 135 studies that examined the mechanisms underlying such effects (see also Madon, Jussim, & Eccles, 1997; Smith et al., 1998). They identified 16 teacher behaviors that are important in the mediation of



positive expectations, including, among others, creating a warmer socioemotional climate, providing more input, having longer interactions, asking more questions, encouraging students, and praising them more. According to the current body of knowledge, teacher misjudgments may very well occur. This is evident, for instance, in the meta-analysis by Südkamp et al. (2012) in which a substantial proportion of the variation of TJs was found to be *unrelated* to actual student performance. (This proportion of variance amounts to approximately $1 - r^2 = 1 - .63^2 = .60$ on average.) However, the effects of SFPs in classroom contexts have typically been found to be small (Jussim & Harber, 2005).

2.3. Open questions

Even after almost half a century of intensive research on TJs (e.g., Rosenthal & Rubin, 1978), major unknowns still exist (Jussim & Harber, 2005): (a) Do TJs affect intelligence? (b) Are there long-term effects of TJs on students' development? (c) Do the effects of TJs generalize across a broader range of key life outcomes?

2.3.1. Do TJs affect intelligence?

First, it is still unclear whether intelligence is actually affected by TJs and the SFPs that potentially result from them (e.g., Rosenthal, 1995; Snow, 1995; Spitz, 1999; Wineburg, 1987). Intelligence results from the interplay of genetic and nongenetic influences (Neisser et al., 1996). Research in the field of intelligence has demonstrated that the (partial) genetic basis of intelligence can be much more easily identified than can concrete environmental influences that have lasting effects on intelligence. Thus, interest was high, but acceptance was mixed (see e.g., Jussim & Harber, 2005, p. 136ff.) when Rosenthal and Jacobson (1968) claimed that, of all things, TJs would represent such a significant nongenetic cause. Depending on how the data from the original Pygmalion study (Rosenthal & Jacobson, 1968) and its intelligence-related follow-up studies (e.g., Rosenthal & Rubin, 1978) are interpreted, one can conclude that either TJs influence intelligence (e.g., Raudenbush, 1984; Rosenthal, 1995) or the exact opposite (e.g., Snow, 1995; Spitz, 1999; Wineburg, 1987). The most persistent points of this issue (for a comprehensive review of the debate, see Jussim & Harber, 2005) are whether or not intelligence in the original Pygmalion study (Rosenthal & Jacobson, 1968) had been adequately operationalized (Snow, 1995) and whether or not the weak-to-nonexistent SFP effects on intelligence in follow-up studies managed to meet the criteria for statistical significance (Wineburg, 1987; see also Raudenbush, 1984).

2.3.2. Are there long-term effects of TJs on student development?

Effects of SFPs on student development seem to dissipate rather than accumulate over time. However, the body of empirical knowledge on which this conclusion is grounded is relatively small and limited to effects on students' educational attainment (Jussim & Harber, 2005). In a sample of some 3000 male high school students, West and Anderson (1976) analyzed how TJs of students' academic performance in Grade 9 were related to students' academic achievement in Grades 10 and 12. TJs from Grade 9 predicted both Grade 10 and Grade 12 *grade point average* (GPA), indicating that TJs can exert effects across time spans that cover several years. Furthermore, considering that TJs had a stronger effect on Grade 10 GPA ($\beta = .12$) compared to Grade 12 GPA ($\beta = .06$), West and Anderson (1976) provided evidence that TJ effects tend to decline rather than accumulate over time. Unfortunately, there was one major shortcoming to the West and Anderson (1976) study. Because their analysis failed to control for students' prior achievement, it remained unclear how accurate the TJs actually were. Hence, their longitudinal results hardly allowed them to make any claims about SFPs.

This was different in a study by Smith, Jussim, and Eccles (1999), who did adjust for students' past performance. Examining 1728 sixth graders, they came to the same conclusion as West and Anderson (1976). Even after controlling for students' prior achievement and other student characteristics, TJs from Grades 6 and 7 predicted significant changes in student achievement, operationalized as school grades and *Michigan Educational Assessment Program* (MEAP) scores, up to Grade 12. As in the West and Anderson (1976) study, predictions were significantly stronger for grade levels closer to the TJs.

Hinnant, O'Brien, and Ghazarian (2009), using a sample of 966 first graders, examined how TJs of children's academic performance in Grade 1 were related to children's academic achievement in Grades 3 and 5. Controlling for differences in prior achievement and additional characteristics of the children, the authors found that TJs of children's math abilities in Grade 1 predicted math achievement in both the third ($\beta = .36$) and fifth grades ($\beta = .09$). In line with West and Anderson (1976) and Smith et al. (1999), Hinnant et al. (2009) found no evidence that SFPs accumulated over time but, once again, they found clear indications for their durability across a long period of time. However, how far this period can extend still remains an open question.

To the best of our knowledge, Alvidrez and Weinstein's (1999) longitudinal study on the effects of TJs extends across the longest period of time so far. Using a sample of 110 four-year-olds, the researchers investigated how TJs of children's intelligence in preschool were related to their high school performance 14 years later. Even after controlling for childhood differences in IQ, *socioeconomic status* (SES), and further childhood characteristics, TJs significantly predicted GPA ($\beta = .37$) and performance on the *Scholastic Aptitude Test* (SAT) taken in late adolescence. Given that the Alvidrez and Weinstein (1999) study included only two points of measurement, it cannot answer whether SFP effects accumulate or dissipate over time. Nevertheless, their results indicate that SFPs can remain effective across a time period of almost 1.5 decades.

2.3.3. Do the effects of TJs generalize across a broader range of key life outcomes?

Much previous research on TJs has centered around the effects of TJs on the development of intelligence. This focus, however, is clearly too narrow if one wants to thoroughly investigate the potentially far-reaching long-term consequences of TJs on students' lives. Certain life outcomes are widely recognized as important, as much for the individuals as for the societies they live in (Ozer & Benet-Martínez, 2006). First and foremost, educational attainment can be considered to be a consequential outcome, as it is a prerequisite for effective and productive participation in modern societies (e.g., OECD, 2010). In line with Ozer and Benet-Martínez (2006), we further assert that most people care about their own health, their well-being, and their socioeconomic situation. If TJs were able to successfully predict such vital outcomes, this would confirm their practical importance from a lifespan perspective and thus significantly enlarge the current body of knowledge in which these outcomes have not yet been included.

Why should TJs predict vital life outcomes? Intelligence is an excellent predictor of many important life outcomes, including educational attainment, health, and socioeconomic status (e.g., Gottfredson & Deary, 2004; Kuncel, Hezlett, & Ones, 2004; Neisser et al., 1996; Zagorsky, 2007). Hence, it is plausible that this is also the

Fig. 1. Path analysis models used to examine the interplay between childhood characteristics at age 12, *teacher judgments* (TJs) at age 12, and life outcomes at age 52. (a) Model 1 examines the potential effects of TJs adjusting for children's IQ at age 12. (b) Model 2 examines the potential effects of TJs adjusting for children's IQ, GPA, gender, parental SES, and nationality. (c) Model 3 examines the potential mediated effects (through years of education) of TJs adjusting for children's IQ, GPA, gender, parental SES, and nationality. Standardized path coefficients that do not involve IQ at age 52 apply likewise to all other life outcomes. Standardized path coefficients representing the relations between childhood characteristics and life outcomes are displayed in Table 2. Intercorrelations between childhood characteristics are shown in Table 1. Coefficients in bold are significant at the .05 level (2-tailed).

case for TJs of student intelligence. In the same vein, given that TJs are believed to affect students' intelligence in a self-fulfilling way (e.g., Rosenthal, 1995; Snow, 1995; Spitz, 1999; Wineburg, 1987), TJs may self-fulfillingly affect life outcomes that are (directly or indirectly) related to intelligence. Importantly, these relations may be mediated via educational attainment because, as is well known, intelligence is positively related to educational attainment (see e.g., Deary, Strand, Smith, & Fernandes, 2007; Kuncel et al., 2004; Neisser et al., 1996). In contrast to the potential effects of SFPs on intelligence, statistical associations of SFPs in the classroom on student achievement have been repeatedly shown and are therefore widely accepted to be real (Jussim & Harber, 2005). Thus, TJs may have long-term effects on students' educational attainment because misjudgments can lead to perceptually biased grading (Jussim, 1989; see also Jussim & Eccles, 1992), and these partially flawed grades can in turn set the course for students' further educational pathways and life chances.

Such effects on educational attainment may have broad and long-term consequences for a broad range of key life outcomes. First, socioeconomic achievement hinges on educational attainment because educational attainment opens the gate to better paying and more prestigious jobs (see e.g., Afarian & Kleiner, 2003; Roth et al., 1996; Roth & Clarke, 1998; Schuler et al., 1990; Strenze, 2007; Zagorsky, 2007). Second, health and *subjective well-being* (SWB) are known to be related to socioeconomic achievement because SES affects access to goods (e.g., housing, nutrition) and services (e.g., health care). Restricted access to such vital resources goes along with an increased exposure to violence and other debilitating and chronic stressors, which in return lead to psychological and physiological distress and ultimately to mental and physical illness (APA Task Force on Socioeconomic Status, 2007; for health, see also Adams, 2002; Gottfredson & Deary, 2004; Johnson, Corley, Starr, & Deary, 2011; for SWB, see also Judge, Ilies, & Dimotakis, 2010; Maruyama, Rubin, & Kingsbury, 1981; Witter, Okun, Stock, & Haring, 1984).

3. The present study

TJs of student intelligence may affect student learning, and they may shape students' educational careers and future social and economic opportunities. However, even after almost half a century of intensive research on TJs of student intelligence, we still do not know whether (a) TJs are associated with the development of intelligence, (b) the relation between TJs and students' development is sustainable, and (c) TJs have (long-term) relations to life outcomes that are directly or indirectly related to student intelligence (see Jussim & Harber, 2005).

The present study aimed to rigorously examine these questions. We therefore drew on data from a prospective cohort study to investigate how TJs of student intelligence at age 12 would predict adult intelligence and key life outcomes (i.e., educational attainment, socioeconomic achievement, SWB, and health) that are likely to be (directly or indirectly) associated with intelligence 40 years later at age 52. To this end, we studied three sets of path analytic models. Basically, each model set addressed all three research questions outlined above. By progressively expanding our model sets (from Model 1 to Model 3; see Sections 3.1 to 3.3), we were able to increasingly provide more conservative estimates of potential TJ effects. In order to strengthen the ecological validity of these findings, we relied on naturalistic data (Alvidrez & Weinstein, 1999). However, given the present research design, hypotheses and research questions were phrased in terms of certain variables predicting rather than causing other variables (Jussim, 1989). Note that in the present study, we used the terms *IQ/intelligence at age 12* and *childhood IQ/intelligence* synonymously. Likewise the terms *IQ/intelligence at age 52* and *adult IQ/intelligence* were used interchangeably.

3.1. Model 1

Model 1 (see Fig. 1a, left side), which capitalizes on the work by Jussim (1989), illustrates the interplay between student intelligence at age 12, TJs of student intelligence, and the intelligence of the same students at age 52. All explanations that are given explicitly for adult intelligence as an outcome apply likewise to all other life outcomes under investigation. In this first model, student intelligence at age 12 is directly related to TJs of these students' intelligence (Path 1). Strong positive Path 1 coefficients indicate that higher levels of student intelligence tend to be associated with higher TJs of these students' intelligence. Hence, Path 1 represents the *accuracy* of TJs. Further, both student intelligence at age 12 and TJs are directly related to intelligence at age 52 as well. This enabled us to study whether intelligence at age 12 explains incremental variance in adult intelligence when TJs are controlled for. Likewise, the path that links TJs to intelligence at age 52 (Path 2) represents whether TJs explain incremental variance in this outcome when childhood intelligence is controlled for. Note that for all three research questions, Path 2—and its derivatives (see next paragraph)—is the primary path of interest.

Importantly, the product of Path 1 and Path 2 represents how the accuracy of TJs of student intelligence is related to adult intelligence. This product term, which we call the *accuracy parameter*, represents an estimate of how adult intelligence is indirectly (and accurately) explained by childhood intelligence through TJs. On the other hand, when this product term is subtracted from the direct path that links TJs to adult intelligence (i.e., Path 2), one obtains an estimate of how adult intelligence depends on TJs above and beyond the accuracy parameter. In other words, this latter estimate, which we call the *expectancy parameter*, represents the potential effect of TJs on adult intelligence that is not attributable to accurately judged individual differences in childhood intelligence.

From a theoretical point of view, the expectancy parameter may be interpreted in two ways. First, the SFP interpretation suggests that TJs predict students' future achievement even after controlling for students' prior achievement (Jussim, 1989). Thus, the expectancy parameter would be an effect attributable to the SFP (Rosenthal & Jacobson, 1968). Second, the expectancy parameter may also reflect an aspect of teachers' diagnostic competence to accurately identify actual preexisting differences among students that are not captured by an intelligence test (Alvidrez & Weinstein, 1999; Jussim, 1991). Both explanations are possible, and neither can be definitively ruled out when using nonexperimental research. We therefore refrain from making claims about whether SFPs actually *cause* differences in students' life outcomes. Note that whenever we speak of *TJ effects* in the present study, we explicitly refer to them as *potential* or *possible* effects because of the above-mentioned uncertainty.

3.2. Model 2

One problem with Model 1 is that TJs may reflect not only student intelligence but also additional childhood characteristics that Model 1 does not take into account. Furthermore, it is always possible that some important third variable, which predicts both TJs and life outcomes, has been omitted (Jussim & Harber, 2005). Thus, per definition, path coefficients in Model 1 represent upper rather than lower-bound estimates (Jussim & Harber, 2005). In order to delimit and isolate the potential effects of TJs adequately, and to anticipate overestimation of the latter at least partially, we integrated a selection of childhood characteristics into *Model 2* (see Fig. 1b, left side), namely, GPA, gender, parental SES, and nationality. These childhood characteristics were selected because they are known (or at least suspected) to influence either the TJ (e.g., Alvidrez & Weinstein, 1999; Dusek & Joseph, 1983; Ready & Wright, 2010; Tenenbaum & Ruck, 2007), the life outcome (e.g., Afarian & Kleiner, 2003; Johnson et al., 2011; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Schuler et al., 1990), or both.

Model 2 is an expansion of Model 1. In addition to the paths already present in Model 1, we introduced children's GPA, gender, parental SES, and nationality into Model 2 as predictors of TJs and life outcomes. Childhood IQ and all childhood characteristics were allowed to be mutually correlated. Strong positive path coefficients linking intelligence at age 12 to TJs (i.e., Path 1) indicate that higher levels of student intelligence tend to be associated with higher TJs of these students' intelligence while controlling for GPA, gender, parental SES, and nationality. Given that in Model 2, the variance in TJs that is explained by childhood intelligence is controlled for the influence of further childhood characteristics, Path 1 represents a more conservative estimate of TJ accuracy than in Model 1. The same logic applies to model paths involving life outcomes. Model 2 allows us to study whether intelligence at age 12 explains incremental variance in life outcomes at age 52 when TJs and further childhood characteristics are controlled for. Likewise, the path that links TJs to life outcomes (i.e., Path 2) now examines whether TJs explain incremental variance in these outcomes when we adjust for childhood IQ and other childhood characteristics. Consequently, in Model 2, both the accuracy parameter and the expectancy parameter are adjusted for children's GPA, gender, parental SES, and nationality.

3.3. Model 3

Model 2 examines potential effects of TJs on life outcomes, accounting for children's IQ and other childhood characteristics. However, considering that TJs influence grading and promotion decisions, it is likely that the potential effects of TJs on life outcomes are mediated via educational attainment, which hinges on these decisions. To explore this mediational path, we proposed *Model 3* (see Fig. 1c, left side). Educational attainment (as a mediator) was operationalized by the number of years of education that a student spent in secondary and tertiary educational training after the completion of primary school. Notably, years of education is a well-established measure of educational attainment (e.g., Neisser et al., 1996) because it reflects exposure to educational opportunities (especially when adjusting for the number of grades a student had to repeat; see Section 4.2.4).

Model 3 is an expansion of Model 2. In addition to the paths already present in Model 2, Model 3 introduces educational attainment as a mediator variable. Model paths involving TJs as a dependent variable (e.g., Path 1) can be interpreted similarly to the corresponding paths in Model 2. The path that directly links TJs to life outcomes (i.e., Path 2) now examines whether TJs explain incremental variance in these outcomes after adjusting for childhood IQ, other childhood characteristics, and children's educational attainment.

4. Method

4.1. Design and sample

The present study analyzed data from the longitudinal study MAGRIP (Brunner & Martin, 2011). MAGRIP used a prospective epidemiological cohort design spanning 40 years. It encompassed two points of measurement: 1968 and 2008.

In 1968, a nationally representative sample of $N = 2824$ Luxembourg students ($M = 11.9$ years, $SD = 0.6$; 50.1% males) participated in a school-based data collection. Every school in Luxembourg participated with data collected from at least one class. The study included a random sample of about half of all Luxembourg Grade 6 classes in 1968. Students filled out a background questionnaire and completed a comprehensive intelligence test, which was administered by trained university students in a group setting. Teachers of participating classes answered a questionnaire in which they reported each student's grades and judged his or her intelligence.

In 2008, a multistep sampling process was applied to derive the sample for the present analyses. First, a random sample of $n = 1632$ persons for whom current addresses could be found was selected to participate in the second wave. This sample was stratified by region of residence in 1968 and gender. Second, of these 1632 persons, 745 people participated in the follow-up sample, 587 persons were contacted but refused to participate, and 300 persons could not be contacted. Trained interviewers visited the study participants of the follow-up study at home and administered various questionnaires. Third, almost half of the participants of the follow-up study ($n = 378$) agreed to retake the same intelligence test they took in 1968. About two thirds ($n = 247$) took the test in a group setting; the remaining participants were visited at home to take the test individually. Analyses for selection bias showed that the follow-up sample was fairly representative of the original sample. Relative to the total 1968 sample, follow-up participants had only slightly higher mean childhood intelligence (Cohen's $d = 0.21$) and SES ($d = 0.08$). Participants who agreed to retake the IQ test at age 52 showed, on average, significantly higher IQ, GPA, and TJs at age 12 as compared to those who refused to be retested. Participants who agreed to retake the IQ test were also found to have significantly higher average educational achievement, SES, and functional health in middle adulthood.

To define the sample of the present study, we applied two criteria in combination. First, we included only participants who took part in the 2008 household study. Second, we included only students for whom a TJ of intelligence was available, as this was the primary variable of interest in the current study. The analyses of the present paper are thus based on data from $n = 731$ persons ($M = 51.8$ years, $SD = 0.6$; 46.5% males). For simplification, we will refer to people at the first wave of measurement as “age 12” and at the second wave of measurement as “age 52” throughout the remaining article.

4.2. Measures

4.2.1. TJs

TJs of student intelligence at age 12 were assessed in the middle of the term by the following question: “How would you rate this child's intelligence?” Responses were given on a 5-point rating scale (1 = *very low*, 2 = *low*, 3 = *average*, 4 = *high*, 5 = *very high*).

4.2.2. Intelligence

Intelligence at ages 12 and 52 was assessed by nine subtests taken from the *Leistungsprüfsystem* (L-P-S, Performance Test System; Horn, 1962), a well-validated standardized (e.g., Sturm & Büssing, 1982) German intelligence test battery. Fluid reasoning, visual processing, and processing speed were each assessed by two subtests. Comprehension knowledge was captured by three subtests. In 1968, the children were randomly administered one of two parallel test forms of the L-P-S. Because the means and variances of the subtests differed slightly across test forms, we used a linear-conversion rule (Kolan & Brennan, 1995) to equate the test scores. To this end, we *z*-standardized the subscales separately for each test form across the entire 1968 sample. In 2008, the participants were given exactly the same test form and items they had completed in 1968. To allow for meaningful comparisons across time, subtest scores from the second wave of measurement were equated using the same conversion rules as had been applied in 1968 (i.e., the standardization of measures in 2008 was based on the means and *SD*s obtained from the entire sample in 1968). For both waves of measurement, all nine subtests were averaged to create a single IQ score, reflecting the person's level of intelligence. For a more convenient interpretation, the resulting composite scores at age 12 were standardized across the entire 1968 sample to the typical IQ metric ($M = 100$, $SD = 15$). Using the same logic, composite scores at age 52 were standardized to $M = 100$ and $SD = 15$. Hence, IQ scores at age 52 were represented on the typical IQ metric of the 12-year-olds.

4.2.3. Childhood characteristics at age 12

Grade point average (GPA) was calculated by using teacher-reported school grades in three subjects (i.e., mathematics achievement, German achievement, and French achievement). Notably, subject-specific grades were taken from three consecutive trimestrial school certificates that preceded the point of testing of the MAGRIP study. All grades were given on a 60-point scale (with higher scores indicating better grades). In general, promotion to subsequent grades required a minimum of 30 points for each subject. To create a single GPA score, we averaged grades across subjects and trimesters.

Children reported their *gender* (dummy-coded as 0 = *female*, 1 = *male* for the statistical analyses) and *nationality* (dummy-coded as 0 = *not Luxembourgish*, 1 = *Luxembourgish*).

Parental SES was assessed through children's indications of their parents' occupations. For the present study, occupations were coded using the *International Standard Classification of Education and Income* (ISEI; Ganzeboom, De Graaf, & Treiman, 1992). The ISEI's theoretical range spans from 16 (farm-hands, laborers, helpers, and cleaners) to 90 (judges), with higher ISEI values indicating higher SES. For the present analyses, we kept the original metric and used only the highest parental ISEI.

4.2.4. Life outcomes at age 52

Except for IQ, all life outcomes were represented by two alternative self-report measures to judge the generalizability of findings.

Educational attainment was operationalized as *years of education*, and whether or not the student had managed to get an *academic degree*. Years of education refer to the number of school years attended after Grade 6. To make sure that the variable reflected actual exposure to educational opportunities rather than time spent in a classroom, repeated grades were counted only once. Academic degree, that is, whether or not the participant successfully completed a tertiary education program (ISCED Level 5 or above; UNESCO, 1997), was introduced as a dummy-coded variable (0 = *no academic degree*, 1 = *academic degree obtained*).

Personal income and *SES* served as measures of socioeconomic achievement. The ISEI of the participant's current or recent occupation (in the case of unemployment at follow-up) served as a measure of the person's SES. Personal income was assessed with a 14-point ordinal scale ranging from *less than 150 €* to *10,000 € or more*. Participants also had the opportunity to report whether they had no personal income. In the present study, we used the mean of each income category divided by 1000 (to allow for a more robust estimation of model parameters; e.g., Kline, 2005) to create a scale ranging from 0 (*no personal income*) to 10 (*10,000 € or more*).

By measuring *satisfaction with life* and *self-esteem*, two important aspects of *subjective well-being* (SWB; Diener, Suh, Lucas, & Smith, 1999) were assessed. Satisfaction with life was measured by the five-item *Satisfaction With Life Scale* (SWLS; Diener, Emmons, Larsen, & Griffin, 1985). Self-esteem was assessed using four items (two inverted) from the Rosenberg (1965) self-esteem scale, with higher scores indicating higher self-esteem. Participants rated their life satisfaction and self-esteem on a 6-point rating scale ranging from 1 = *strongly disagree* to 6 = *strongly agree*. For both scales, the total sum score across all items was computed in terms of the *Percent of Maximum Possible* (POMP) score,[1] with 100% indicating the maximum and 0% the minimum possible score.

Finally, we analyzed *subjective* and *functional health*, which represent two important dimensions of adult health (Liang, 1986). Subjective health was assessed by participants' ratings of their health status overall and relative to their peers (5-point scale) and their satisfaction with their health status (7-point scale). These items have been shown to measure distinct aspects of subjective health and have been used in influential studies and large-scale surveys (Liang, 1986; Terwey, 2000). Functional health was assessed by the following question: "Looking back over the past 2 years, how much did your health status hinder the following activities?" Nine areas of activity were listed (e.g., work, household, sports). Responses were given on a 5-point rating scale, with higher scores indicating greater impairment; for statistical analyses, all items were inverted. For both subjective and functional health, total POMP scores were computed as described above.

Descriptive statistics (including scale score reliabilities) and intercorrelations for all applied measures are listed in Table 1.

4.3. Statistical analysis

Fig. 1 depicts the path analysis Models 1 to 3 that we used to investigate our research questions. All analyses were performed with the statistical software package Mplus (L.K. Muthén & Muthén, 2010). Path coefficients reported in this article may be interpreted as standardized betas from a regression equation. Note that all models that we analyzed are so-called just-identified models with zero degrees of freedom. An intrinsic property of these models is that they demonstrate perfect fit to the empirical data. Analogous to multiple regression models, such just-identified models are highly informative concerning the strength of the relations between the constructs under investigation because global or local model misfit cannot disturb the estimates of the model parameters (see Tomarken & Waller, 2003).

The *accuracy parameter* (i.e., the potential indirect effect of childhood IQ on life outcome through TJ; see Section 3.1) was computed as $\beta_{TJ\ on\ IQ} \times \beta_{Life\ outcome\ on\ TJ}$. The *expectancy parameter* (i.e., the potential incremental effect of TJ on life outcome that is not attributable to the differences in childhood IQ; see Section 3.1) was computed as $\beta_{Life\ outcome\ on\ TJ} - \beta_{TJ\ on\ IQ} \times \beta_{Life\ outcome\ on\ TJ}$ (see Table 2).

Missing data are an unfortunate reality in large-scale assessment, and our study is no exception. Information regarding the percentages of missing data for each variable can be found in Table 1. When IQ at age 52 was excluded, the amount of missing data per measure ranged between 0% and 9%. The percentage of missing data on IQ at age 52 (51%) was considerably higher because only half of the participants agreed to retake the intelligence test. To efficiently deal with the pattern of missing data that we observed in the present study, we applied the *Full Information Maximum Likelihood* (FIML) estimation (B.O. Muthén, Kaplan, & Hollis, 1987). To further improve the quality of the parameter estimates, we used all variables from Table 1 (except those that were included in the corresponding model) as *auxiliary* variables (see Graham, 2003; L.K. Muthén & Muthén, 2010).

5. Results

Before we turn to the results obtained for Models 1 to 3, we will present the results that we obtained when we examined the predictive power of TJs for life outcomes. The highest correlation was observed between TJs of student intelligence and IQ at age 52 (r = .50; see Table 1). Correlations between TJs and indicators of educational attainment were medium to large in size. Quite a similar pattern was found for socioeconomic achievement. Furthermore, TJs showed small correlations with both health scales but failed to correlate significantly with measures of SWB. Table 1 juxtaposes how the life outcomes that were related to TJs were also related to childhood intelligence and parental SES, which are both well-established long-term predictors of many life outcomes (e.g., Kuncel et al., 2004). As Table 1 clearly illustrates, TJs showed stronger relations with these outcomes than parental SES. Likewise, (except for IQ at age 52), the majority of outcomes were more highly correlated with TJs than with childhood IQ.

[1] POMP = [(observed − minimum)/(maximum − minimum)] × 100, where observed = the observed score for a single case, minimum = the minimum possible score on the scale, and maximum = the maximum possible score on the scale (Cohen, Cohen, Aiken, & West, 1999, p. 323).

Table 1
Intercorrelations between childhood characteristics at age 12 and life outcomes at age 52.

r	Age 12						Age 52								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Teacher judgment	–														
2 IQ	**.51**	–													
3 GPA	**.79**	**.51**	–												
4 Gender	−.06	.03	**−.13**	–											
5 Parental SES	**.25**	**.18**	**.28**	−.03	–										
6 Nationality	.04	**.16**	**.11**	−.01	**.19**	–									
7 IQ	**.50**	**.75**	**.52**	.03	**.23**	**.11**	–								
8 Years of education	**.44**	**.38**	**.46**	**.18**	**.36**	**.07**	**.47**	–							
9 Academic degree	**.34**	**.30**	**.35**	**.10**	**.30**	.07	**.31**	**.70**	–						
10 SES	**.46**	**.34**	**.49**	.00	**.29**	.02	**.37**	**.56**	**.48**	–					
11 Income	**.19**	**.23**	**.17**	**.53**	**.14**	.05	**.26**	**.46**	**.33**	**.38**	–				
12 Satisfaction with life	.02	.01	.01	**.08**	.02	.06	.04	.07	**.08**	.06	**.14**	–			
13 Self-esteem	.07	**.09**	**.09**	**.08**	−.03	.00	**.11**	.05	.04	**.09**	**.19**	**.40**	–		
14 Subjective health	**.12**	.06	**.09**	−.05	**.08**	.02	**.15**	**.13**	**.08**	**.15**	.06	**.37**	**.20**	–	
15 Functional health	**.18**	**.13**	**.16**	.03	**.08**	.01	**.21**	**.20**	**.16**	**.22**	**.13**	**.24**	**.08**	**.56**	–
Mean	3.50	102.20	45.86	.47	40.24	.86	126.76	5.84	.15	46.73	3.17	69.25	75.19	68.76	82.74
SD	1.03	8.95	8.55	.50	12.68	.34	14.27	3.44	.35	14.55	2.36	16.66	14.37	22.82	20.24
Cronbach's alpha	–	.80	.97	–	–	–	.86	–	–	–	–	.83	.62	.89	.93
Percentage of missing data	.00	.00	.08	.00	.00	.01	.51	.00	.01	.00	.08	.03	.02	.02	.09

Note. Correlations, means, and standard deviations were computed with the full information maximum likelihood estimator. Gender is a dummy-coded variable (0 = female, 1 = male). Nationality is a dummy-coded variable (0 = other, 1 = Luxembourgish). Academic degree is a dummy-coded variable (0 = No, 1 = Yes). Coefficients in bold are significant at the .05 level (2-tailed).

5.1. Model 1

Fig. 1 (right side) provides a detailed overview of all model paths exemplified for IQ at age 52 as outcome variables. Model paths that do not involve IQ at age 52 remain identical for all life outcomes. Model paths that do involve IQ at age 52 change with the outcome under investigation and are therefore reported separately for each life outcome in Table 2. *Model 1* revealed that childhood IQ was a strong predictor of TJs (β = .51; see Fig. 1a, right side). However, Model 1 explained only about a quarter of the variance (R^2 = .26) that occurred in these judgments. Most importantly, Model 1 confirmed that the relations between TJs and life outcomes (as observed in Table 1) remained even after controlling for childhood IQ (see Table 2). TJ predictions were strongest for academic degree (β = .41), followed closely by SES (β = .38). Satisfaction with life and self-esteem were not influenced by TJs or by IQ at age 12. Childhood IQ predicted neither of the two health scales, whereas TJs were a small yet significant predictor of both. In general, the relations between life outcomes and TJs were attributable in about equal amounts to accuracy and expectancy parameters. Most notably, TJs significantly predicted adult IQ above and beyond childhood IQ: The direct path (i.e., Path 2) was β = .16, and the accuracy parameter and expectancy parameter both amounted to β = .08, respectively.

5.2. Model 2

When we integrated further childhood characteristics[2] into *Model 2*, the pattern of results changed considerably. Although childhood IQ was still a significant predictor of TJs, its power to predict TJs had weakened substantially (β = .15; see Fig. 1b, right side). By contrast, GPA was found to be a strong predictor of TJs (β = .72). Consequently, Model 2 explained about two thirds of the variance in TJs (R^2 = .65). With the introduction of control variables, the predictive power of TJs dropped and even lost significance for the prediction of IQ and academic degree (see Table 2). GPA seemed to compensate for this loss. This was particularly the case for academic degree (β = .46), but even IQ at age 52 was influenced by GPA (β = .14). Not only did the introduction of GPA affect the magnitude of the direct predictive power of TJs, but it also affected the ratio of accuracy parameters to expectancy parameters in these predictions. Given that childhood IQ was no longer the strongest predictor of TJs in Model 2, the ability of TJs to predict life outcomes was revealed to be less accurate, resulting in smaller accuracy parameters compared to expectancy parameters (see Table 2).

5.3. Model 3

As we already knew from Model 2, years of education was significantly influenced by all age-12 variables except for nationality (see Table 2). The results from *Model 3* revealed that once years of education had been interposed between TJs and life outcomes, years of education almost completely mediated the potential effects of TJs (and childhood characteristics) on these outcomes (see Fig. 1c, right side). All life outcomes except for measures of SWB were influenced by years of education (see Table 2). As was expected, years of education was a particularly strong predictor of academic degree (β = .75), whose variance was now mostly explained (R^2 = .82). Direct paths relating TJs to life outcomes were at best small (for SES: β = .09; for both health dimensions: βs = .11) and did not reach statistical significance.

6. Discussion

TJs of student intelligence may affect students' learning, thus shaping their educational careers and future social and economic opportunities. However, even almost half a century of intensive research on TJs of student intelligence had still not revealed whether (a) TJs are associated with the development of intelligence, (b) the relations between TJs and students' development are sustainable, and (c) TJs have (long-term) relations to life outcomes that are directly or indirectly related to student intelligence (see Jussim & Harber, 2005). The present paper aimed to rigorously examine these questions.

The main results of the present study can be summarized as follows: First, TJs of intelligence reflected GPA rather than IQ; this is important to keep in mind for the interpretation of all further findings. Second, TJs

[2] A striking additional finding of the present study was a strong gender effect observed for income at age 52 (β = .55; see Table 2). When all other variables (childhood IQ, GPA, parental SES, and nationality) were held constant, men earned on average 2593 € net per month more than women. The large gender difference in income also remained at 1893 € in favor of men even when we controlled for whether a person was unemployed or not. Notably, the size of this (adjusted) gender difference in income is comparable in size to the difference in income between a person who left school after primary school and a person who graduated from high school (i.e., 1617 €).

Table 2
Structural relations between childhood characteristics at age 12, teacher judgments at age 12, and life outcomes at age 52.

Life outcome at age 52	Model	IQ		Teacher judgment						Years of education		GPA		Gender		Parental SES		Nationality		$R^2_{\text{Life outcome}}$
				Direct path		Accuracy parameter		Expectancy parameter												
		β	p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	
IQ																				
	1	**.67**	.00	**.16**	.00	**.08**	.00	**.08**	.00											**.59**
	2	**.65**	.00	.05	.41	.01	.42	.04	.42			**.14**	.03	.03	.35	.06	.06	−.02	.49	**.60**
	3	**.62**	.00	.03	.60	.01	.60	.03	.60	**.16**	.00	.10	.13	.00	.94	.02	.50	−.02	.57	**.61**
Years of education																				
	1	**.22**	.00	**.33**	.00	**.17**	.00	**.16**	.00											**.23**
	2	**.15**	.00	**.11**	.04	**.02**	.05	**.09**	.04			**.27**	.00	**.22**	.00	**.24**	.00	−.03	.37	**.35**
	3	–		–		–		–		–		–		–		–		–		–
Academic degree																				
	1	**.26**	.00	**.41**	.00	**.21**	.00	**.20**	.00											**.35**
	2	**.15**	.01	.09	.28	.01	.28	.08	.28			**.46**	.00	**.21**	.00	**.23**	.00	−.03	.64	**.55**
	3	.02	.70	.11	.18	.02	.18	.10	.18	**.75**	.00	.11	.31	.03	.45	.05	.23	−.01	.79	**.82**
SES																				
	1	**.14**	.00	**.38**	.00	**.20**	.00	**.19**	.00											**.22**
	2	**.09**	.01	**.13**	.01	**.02**	.03	**.11**	.01			**.30**	.00	.05	.12	**.17**	.00	**−.07**	.04	**.29**
	3	.04	.33	.09	.07	.01	.09	.08	.07	**.40**	.00	**.19**	.00	−.04	.21	**.08**	.02	−.05	.07	**.39**
Income																				
	1	**.18**	.00	**.10**	.02	**.05**	.02	**.05**	.02											**.06**
	2	**.11**	.00	.07	.21	.01	.22	.06	.21			.10	.06	**.55**	.00	**.09**	.00	.01	.86	**.36**
	3	.06	.09	.03	.54	.00	.54	.03	.54	**.33**	.00	.01	.79	**.48**	.00	.01	.68	.02	.63	**.43**
Satisfaction with life																				
	1	.00	.92	.03	.57	.01	.57	.01	.57											.00
	2	−.02	.63	.04	.55	.01	.56	.03	.55			.00	.95	**.08**	.03	.01	.84	.06	.14	.01
	3	−.03	.50	.03	.62	.01	.62	.03	.62	.06	.19	−.02	.77	.07	.07	−.01	.87	.06	.13	.01
Self-esteem																				
	1	.07	.09	.03	.46	.02	.46	.02	.46											.01
	2	.06	.22	−.02	.71	.00	.71	−.02	.71			.11	.10	**.09**	.02	−.06	.15	−.01	.80	**.02**
	3	.06	.21	−.02	.72	.00	.72	−.02	.72	−.01	.87	.11	.10	**.09**	.02	−.06	.18	−.01	.79	**.02**
Subjective health																				
	1	.01	.91	**.12**	.01	**.06**	.01	**.06**	.01											.01
	2	.01	.82	**.13**	.05	.02	.06	**.11**	.05			−.04	.56	−.05	.23	.05	.17	.00	.94	.02
	3	−.01	.88	.11	.07	.02	.09	.10	.07	**.12**	.01	−.07	.30	−.07	.07	.03	.54	.01	.87	**.03**
Functional health																				
	1	.05	.25	**.16**	.00	**.08**	.00	**.08**	.00											**.03**
	2	.04	.35	**.13**	.05	.02	.06	**.11**	.05			.03	.63	.04	.34	.04	.38	−.01	.74	**.04**
	3	.02	.65	.11	.08	.02	.10	.10	.08	**.15**	.00	−.01	.92	.01	.90	.00	1	−.01	.81	**.05**

Note. Table shows the standardized model parameters (β). Direct path = Path 2 (see Fig. 1). Accuracy parameter = potential indirect effect of childhood IQ on life outcome through teacher judgment. Expectancy parameter = potential incremental effect of teacher judgment on life outcome that is not attributable to the differences in childhood IQ. Gender is a dummy-coded variable (0 = female, 1 = male). Nationality is a dummy-coded variable (0 = other, 1 = Luxembourgish). Academic degree is a dummy-coded variable (0 = No, 1 = Yes). Coefficients in bold are significant at the .05 level (2-tailed). $n = 731$.

predicted important life outcomes (including intelligence) across a time span of 40 years. Third, the predictive power of TJs remained even after controlling for childhood intelligence. TJ predictions were split about evenly into accuracy and expectancy parameters (i.e., potential SFPs). Fourth, when adjusting for GPA, the predictive power of TJs dropped considerably. This emphasized that the predictive power of TJs hearkened back to the fact that TJs mainly reflected students' educational achievement in terms of GPA. Fifth, the potential impact of TJs on life outcomes was almost completely mediated by educational attainment.

6.1. TJ accuracy

Before we tackle the unresolved questions, we will first examine the accuracy of TJs because all further discussion will hinge on this information. Given the positive and fairly high correlation between TJs and childhood IQ, one could easily jump to the conclusion that judgment accuracy, as defined by Südkamp et al. (2012), is assured. However, when analyzed in a path analytic framework, the data revealed that TJs of intelligence reflected GPA rather than IQ. School grades as represented by students' GPAs are a commonly used index of academic achievement; nonetheless, they are a form of TJ themselves (Alvidrez & Weinstein, 1999). The fact that teachers perceived student intelligence more consistently with their own judgments (or the grading judgments of the teachers who taught these students in the previous grade) suggests that the TJs in the present study were perceptually biased (Jussim, 1989; see also Jussim & Eccles, 1992). Although on average, school grades correlate with measures of intelligence at about $r = .50$ (Neisser et al., 1996; $r = .51$ in our study), neither variable can replace the other because they share only about 25% of their variance. It appears, though, that the teachers in our study did exactly that: With $\beta = .15$ (TJs regressed on childhood IQ; see Fig. 1b, right side), the data clearly showed that teachers were hardly able to accurately judge their students' intelligence above and beyond the part of intelligence that was shared by IQ scores and GPA. What are the implications of this finding for the further interpretation of the results? In the following discussion, one has to keep in mind that, although teachers were explicitly asked to judge their students' intelligence, they mostly provided a very accurate assessment ($r = .79$) of students' educational achievement in terms of GPA.

6.2. Do TJs affect intelligence?

Even after almost half a century of intensive research on TJs of students' intelligence (e.g., Rosenthal & Rubin, 1978), the question of whether intelligence is affected by TJs and the SFPs that result from them (see e.g., Rosenthal, 1995; Snow, 1995; Spitz, 1999; Wineburg, 1987) is still an ongoing controversy (Jussim & Harber, 2005). In the present study, of all life outcomes under investigation, IQ at age 52 was the one that was most highly correlated ($r = .50$) with TJs, indicating that TJs were long-term predictors of student intelligence. With a statistically significant TJ expectancy parameter of $\beta = .08$ in Model 1, the present study further provided evidence for possible effects of SFPs on intelligence. However, with the expansion of our models, it became clear that it was years of education, and not TJs, that significantly predicted—and potentially influenced—adult IQ above and beyond childhood IQ. This makes sense given that formal schooling and exposure to learning opportunities are indeed known to enhance intelligence (Ceci, 1991; Gustafsson, 2001; see also Neisser et al., 1996). Nonetheless, the absence of statistically significant direct paths relating TJs to life outcomes in Model 3 did not automatically exclude long-term effects of SFPs on intelligence. One should not forget that, even after controlling for childhood characteristics in Model 2, TJs still had a statistically significant expectancy parameter for years of education as an outcome ($\beta = .09$). Thus, even in our most restrictive Model 3, the possibility that TJs at age 12 self-fulfillingly affect IQ at age 52 through years of education cannot be excluded. However, given the naturalistic nature of our study, we cannot confirm these SFPs for sure (see Section 3.1). What we can confirm, though, is Jussim and Harber's (2005) conclusion that SFPs, if they occur at all, are rather small indeed (which, by the way, was true for all life outcomes under investigation).

6.3. Are there long-term effects of TJs on student development?

Besides the general controversy regarding whether or not SFPs affect intelligence, the sustainability of such effects is of major interest as well (Jussim & Harber, 2005). According to the current body of knowledge, SFPs seem to dissipate rather than accumulate across time (Hinnant et al., 2009; Smith et al., 1999; West & Anderson, 1976). Although there are data that indicate that SFPs can remain effective across a time span of more than a decade (Alvidrez & Weinstein, 1999), prior research has not taken a lifespan perspective. Unfortunately, given that our study encompassed only two points of measurement, we were unable to analyze how (potential) TJ effects can evolve over time. What we can say, though, is that our study provides evidence for sustainable TJ expectancy parameters for IQ as an outcome (mediated through years of education) across a time span of 4 decades. Consequently, the present study considerably widens the body of knowledge on the durability of TJs. Given that all potential SFPs were mediated through educational attainment (which hardly changes after age 52), it is likely that the observed relations will persist for a lifetime.

6.4. Do the effects of TJs generalize across a broader range of key life outcomes?

Finally, it remains an open question whether classroom SFPs have far-reaching, long-term consequences for vital aspects of students' adult lives. In the present study, TJs were equally or more highly related to the broad range of investigated life outcomes than were childhood intelligence and children's parental SES, indicating that future research should consider TJs to be a serious alternative to other well-established long-term predictors of these life outcomes. From a path-analytic perspective, TJs were found to be potentially self-fulfillingly related to measures of educational attainment, socioeconomic achievement, and health. However, as hypothesized (see, e.g., Adams, 2002; Deary et al., 2007; Gottfredson & Deary, 2004; Johnson et al., 2011; Roth et al., 1996; Roth & Clarke, 1998), once again, all potential effects of SFPs on these outcomes were mediated through years of education. Measures of SWB, for their part, were not affected by TJs or by years of education. Hence, although TJs at age 12 may have affected students' educational pathways and thus their cognitive development, their socioeconomic achievement, and even their health in a self-fulfilling way, TJs had no effect whatsoever on the students' satisfaction with themselves and their lives 40 years later.

If we assume that the observed expectancy parameters were SFPs indeed, the good news is that (as all TJ expectancy parameters were strictly positive) an overestimation of student intelligence does no harm. In the best-case scenario, an overestimation may literally pay out (see income in Table 2). But why is that so? Overestimating intelligence could be helpful because teachers' beliefs in what children may be able to accomplish might in turn influence children's beliefs in themselves. Thus, through overestimation, teachers may establish what is known as a *Zone of Proximal Development* (Vygotsky, 1978), supporting actual development in the direction of this goal.

6.5. Strengths and limitations

We utilized representative data to examine the predictive power of TJs for intelligence and additional key life outcomes across a time span of 4 decades. Despite these strengths, our study has several limitations.

First, given the naturalistic design of our study, we could not definitively confirm the effects of SFPs on life outcomes. However, by controlling for numerous alternative explanations, and by filtering out obvious parts of predictive accuracy (i.e., the accuracy parameter), we did our best to narrow down the interpretation of the reported TJ expectancy parameters.

Second, teachers may have taken into account facets of intelligence (e.g., creativity) that the IQ test did not cover. If so, TJs may be more accurate than our statistical models could possibly reveal. Closely related to this second point is another question that future research might address: Are teachers actually competent enough to distinguish between intelligence and educational performance?

Third, and more generally, the potential effects of TJs on life outcomes may not necessarily be attributed to SFPs. TJs of student intelligence could easily—and plausibly—also reflect estimates of conscientiousness, integrity, achievement motivation, and/or further systematic components in TJs that may affect outcomes. The fact that, in the present study, TJs of student intelligence reflected GPA rather than IQ provides possible evidence in this direction, as school grades are known to reflect not only achievement, but also achievement motivation (see e.g., Robbins et al., 2004).

Fourth, in the present paper, we examined the potential long-term effects of TJs on life outcomes by means of path models. The path coefficients as reported in Table 2 do not differentiate effects due to over- or underestimation of students' intelligence. Hence, we could not provide a satisfactory answer to the open question of whether SFPs in general do more harm than good (see Jussim & Harber, 2005).

Fifth, given that our data encompassed two points of measurement only, we were unable to analyze how exactly TJ effects evolve over time. Although the current study brought insights into the as-yet-unexplored long-term impacts of TJs, further research on this topic is definitely needed. In particular, studies including not only a longitudinal perspective and a broad selection of outcomes, but also multiple points of measurement would enlarge the current body of knowledge in a meaningful way. To further scrutinize the ecological validity of TJ effects, future research should also control for plausible student and teacher characteristics. As our Model 3 revealed, potential TJ effects, as hypothesized, were almost entirely mediated by years of education. This finding is new and needs cross-cultural replication to determine whether it generalizes to other educational systems or whether it is specific to Luxembourg.

Sixth, although our hypotheses about why TJs may affect vital life outcomes are reasonably well-founded (see Section 2.3.3), the analysis of outcomes other than intelligence or educational attainment remains —without dispute—exploratory at least to some degree.

7. Conclusions

The present prospective study aimed to clarify major unknowns in the field of TJ research (see Jussim & Harber, 2005) by taking a lifespan perspective. In conclusion, our data supported potential small long-term effects of classroom SFPs on intelligence that were completely mediated by educational attainment. Further, our data supported potential long-term effects of SFPs on other vital life outcomes, including socioeconomic achievement and health, which were also mediated via educational attainment. TJs were not sustainably related to SWB. Given that teachers may self-fulfillingly affect the lives of their students up to middle adulthood and probably even beyond, we believe that it is essential that a teacher's ability to (accurately) assess student achievement (Ready & Wright, 2010) be considered a key aspect of teachers' professional training. Specifically, future teachers should be trained to become reflective (Schön, 1983) but also responsible practitioners who are now more than ever sensitive to the key role their judgments might play in their students' later lives. We hope that the current study can contribute to this crucial purpose.

References

Adams, S. J. (2002). Educational attainment and health: Evidence from a sample of older adults. *Education Economics*, *10*(1), 97–109. http://dx.doi.org/10.1080/09645290110110227.

Afarian, R., & Kleiner, B. H. (2003). The relationship between grades and career success. *Management Research News*, *26*(2/3/4), 42–51. http://dx.doi.org/10.1108/01409170310783781.

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology*, *91*(4), 731–746.

APA Task Force on Socioeconomic Status (2007). *Report of the APA task force on socioeconomic status*. Washington, DC: APA.

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*(7), 631–661.

Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2013). The diagnostic skills of mathematics teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 229–248). New York: Springer.

Brunner, M., & Martin, R. (Eds.). (2011). Die MAGRIP-Studie (1968-2009). Wie beeinflussen sozio-kognitive Merkmale von Kindern im Grundschulalter und ihre Bildungswege ihr späteres Leben als Erwachsene in Luxemburg?. [The MAGRIP study (1968-2009). How do socio-cognitive attributes and educational pathways of primary school pupils shape their adult lives in Luxembourg?]. Luxembourg: University of Luxembourg, EMACS.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*(5), 703–722. http://dx.doi.org/10.1037/0012-1649.27.5.703.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315–346.

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

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. http://dx.doi.org/10.1207/s15327752jpa4901_13.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. http://dx.doi.org/10.1037/0033-2909.125.2.276.

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, *75*(3), 327–346.

Ganzeboom, H. B. G., De Graaf, P.M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*, 1–56.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1–4.

Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 80–100.

Gustafsson, J. -E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal*, *2*(4), 166–186.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, *97*(3), 363–386.

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, *101*(3), 662–670. http://dx.doi.org/10.1037/a0014306.

Horn, W. (1962). *Leistungsprüfsystem (L-P-S) [Performance Test System (L-P-S)]*. Göttingen: Hogrefe.

Johnson, W., Corley, J., Starr, J. M., & Deary, I. J. (2011). Psychological and physical health at age 70 in the Lothian birth cohort 1936: Links with early life IQ, SES, and current cognitive function and neighborhood environment. *Health Psychology*, *30*(1), 1–11. http://dx.doi.org/10.1037/a0021834.

Judge, T. A., Ilies, R., & Dimotakis, N. (2010). Are health and happiness the product of wisdom? The relationship of general mental ability to educational and occupational attainment, health, and well-being. *Journal of Applied Psychology*, *95*(3), 454–468. http://dx.doi.org/10.1037/a0019084.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, *57*(3), 469–480.

Jussim, L. (1991). Social perception and social reality: A reflection–construction model. *Psychological Review*, *98*(1), 54–73. http://dx.doi.org/10.1037/0033-295X.98.1.54.

Jussim, L., & Eccles, J. S. (1992). Teacher expectations: II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, *63*(6), 947–961. http://dx.doi.org/10.1037/0022-3514.63.6.947.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. http://dx.doi.org/10.1207/s15327957pspr0902_3.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.)New York: Guilford Press.

Kolan, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161.

Liang, J. (1986). Self-reported physical health among aged adults. *Journal of Gerontology*, *41*, 248–260.

Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, *72*(4), 791–809.

Maruyama, G. M., Rubin, R. A., & Kingsbury, G. G. (1981). Self-esteem and educational achievement: Independent constructs with a common cause? *Journal of Personality and Social Psychology*, *40*(5), 962–975.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, *8*(2), 193–210.

Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus. Statistical analysis with latent variables. User's guide* (6th ed.)Los Angeles: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101.

OECD (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics and science*, *vol.1*, Paris: OECD Publishing.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*(1), 401–421. http://dx.doi.org/10.1146/annurev.psych.57.102904.190127.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, *76*(1), 85–97.

Ready, D.D., & Wright, D. L. (2010). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335–360. http://dx.doi.org/10.3102/0002831210374874.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. http://dx.doi.org/10.1037/0033-2909.130.2.261.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. http://dx.doi.org/10.1111/j.1745-6916.2007.00047.x.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.

Rosenthal, R. (1995). Critiquing Pygmalion: A 25-year perspective. *Current Directions in Psychological Science*, *4*(6), 171–172. http://dx.doi.org/10.1111/1467-8721.ep10772607.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377. http://dx.doi.org/10.1017/S0140525X00075506.

Roth, P. L., BeVier, C. A., Switzer, F. S., III, & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, *81*(5), 548–556.

Roth, P. L., & Clarke, R. L. (1998). Meta-analyzing the relation between grades and salary. *Journal of Vocational Behavior*, *53*, 386–400.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. London: Temple Smith.

Schuler, H., Funke, U., & Baron-Boldt, J. (1990). Predictive validity of school grades. A meta-analysis. *Applied Psychology: An International Review*, *39*(1), 89–103.

Smith, A. E., Jussim, L., & Eccles, J. (1999). Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *Journal of Personality and Social Psychology*, *77*(3), 548–565. http://dx.doi.org/10.1037/0022-3514.77.3.548.

Smith, A. E., Jussim, L., Eccles, J., Van Noy, M., Madon, S., & Palumbo, P. (1998). Self-fulfilling prophecies, perceptual biases, and accuracy at the individual and group levels. *Journal of Experimental Social Psychology*, *34*(6), 530–561.

Snow, R. E. (1995). Pygmalion and intelligence? *Current Directions in Psychological Science*, *4*(6), 169–171. http://dx.doi.org/10.1111/1467-8721.ep10772605.

Spitz, H. H. (1999). Beleaguered Pygmalion: A history of the controversy over claims that teacher expectancy raises intelligence. *Intelligence*, *27*(3), 199–234. http://dx.doi.org/10.1016/S0160-2896(99)00026-4.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426. http://dx.doi.org/10.1016/j.intell.2006.09.004.

Sturm, W., & Büssing, A. (1982). Ein Vergleich von HAWIE und LPS bei der psychometrischen Einzelfalldiagnostik neurologischer Patienten [A comparison of HAWIE and LPS for psychometric individual case diagnosis of neurological patients]. *Diagnostica*, *28*, 348–359.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/a0027627.

Tenenbaum, H. R., & Ruck, M.D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, *99*(2), 253–273. http://dx.doi.org/10.1037/0022-0663.99.2.253.

Terwey, M. (2000). ALLBUS: A German general social survey. Schmollers Jahrbuch. *Journal of Applied Social Science Studies*, *120*, 151–158.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578–598. http://dx.doi.org/10.1037/0021-843X.112.4.578.

UNESCO (1997). International standard classification of education. Retrieved from. http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes.* Cambridge: Harvard University Press.

West, C. K., & Anderson, T. H. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research*, *46*(4), 613–630.

Wineburg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy. *Educational Researcher*, *16*(9), 28–37.

Witter, R. A., Okun, M.A., Stock, W. A., & Haring, M. J. (1984). Education and subjective well-being: A meta-analysis. *Educational Evaluation and Policy Analysis*, *6*(2), 165–173.

Zagorsky, J. L. (2007). Do you have to be smart to be rich? The impact of IQ on wealth, income and financial distress. *Intelligence*, *35*(5), 489–501. http://dx.doi.org/10.1016/j.intell.2007.02.003.