Social Desirability Bias in Child-Report Social Well-Being: Evaluation of the Children's Social Desirability Short Scale Using Item Response Theory and Examination of Its Impact on Self-Report Family and Peer Relationships

Anne-Linda Camerini & Peter J. Schulz

Child Indicators Research
The official Journal of the International Society for Child Indicators

ISSN 1874-897X
Volume 11
Number 4

Child Ind Res (2018) 11:1159-1174
DOI 10.1007/s12187-017-9472-9

Springer

Child Ind Res (2018) 11:1159–1174
DOI 10.1007/s12187-017-9472-9

Social Desirability Bias in Child-Report Social Well-Being: Evaluation of the Children's Social Desirability Short Scale Using Item Response Theory and Examination of Its Impact on Self-Report Family and Peer Relationships

Anne-Linda Camerini[1] · **Peter J. Schulz**[1]

Accepted: 27 April 2017 / Published online: 12 May 2017

Abstract Research on child well-being largely relies on children's self-report data, potentially biased by social desirability (SD). In this study, we aim to (1) evaluate the psychometric properties of the Children's Social Desirability Short (CSD-S) scale, and (2) examine if and, if so, how SD systematically biases child-report family and peer relationships as indicators of social well-being. In spring 2015, 843 elementary school children (aged 10) and their parents were surveyed on well-being indicators and SD measured with the 14-items Children's Social Desirability Short (CSD-S) scale. The CSD-S was evaluated using nonparametric Item Response Theory (NIRT). Linear mixed-effects regression models based on multiple imputations of multilevel missing data were run to examine the role of SD in self-report social well-being in addition to socio-demographic characteristics, accounting for the nested structure of the data (students were sampled at class level). Applying NIRT, we identified a 13-items subset of the CSD-S with double monotonicity. Cronbach's alpha was .82. When controlling for children's socio-demographic characteristics, SD significantly positively predicted subjective evaluations of family relationships ($B = 0.04$, $t(49272) = 7.45$, $p < .001$), whereas it significantly negatively predicted self-report deviant behavior performed towards peers ($B = -0.03$, $t(39927) = -14.40$, $p < .001$) and experienced from peers ($B = -.0.01$, $t(39028) = -2.86$, $p = .002$). SD bias explained additional 22 percent of variance in self-report deviant behavior performed towards peers. Since SD impacts the

✉ Anne-Linda Camerini
anne.linda.camerini@usi.ch

Peter J. Schulz
peter.schulz@usi.ch

[1] Institute of Communication and Health, Department of Communication Sciences, Università della Svizzera italiana, Via Giuseppe Buffi 13, 6904 Lugano, Switzerland

validity of self-report well-being, child indicators research should include age-specific SD scales, e.g., the CSD-S, and control for the bias in statistical analyses.

Keywords Social desirability bias; social well-being · Item response theory · Italian children's social desirability short scale

1 Introduction

Child well-being has been the focus of increasing scholarly attention as a meaningful public health outcome (Bradshaw et al. 2006; Earls and Carlson 2001). Given the plethora of indicators of child well-being (Hauser et al. 1997), Pollard and Lee (2003) propose a review-based multi-dimensional concept which encompasses *physical* (e.g., healthy body weight, absence of diseases), *psychological* (e.g., absence of mental illnesses, stress), *cognitive* (e.g., cognitive abilities, academic achievement), *social* (e.g., family structure, relationship with peers), and *economic* (e.g., family's economic status) well-being. While some indicators of child well-being, e.g., healthy body weight, can be obtained through non-reactive measures, others like family structure or stress are commonly assessed through self-report. In addition, financial and time resources may force researchers to rely on self-report even when non-reactive methods are at hand.

When working with self-report data, researchers are faced with potential measurement error (Novick 1966), which can be divided into two types: random and systematic error. While random errors sum up to zero across a sufficient number of observations and, thus, can be ignored, systematic error consistently occurs across a number of observations and thereby introduces "bias" in the data. A well-known bias in social sciences is the social desirability (SD) bias (Krumpal 2013), which has been researched since the 1950s with Edwards (1957) and Crowne and Marlowe (1960; Marlowe and Crowne 1961) among the pioneers in this field.

The SD bias describes the tendency of respondents to inaccurately report on sensitive topics in order to present oneself in the best possible light (Ford and Rubin 1970). Dependent on the researched topic, the bias can be either reflected in under-reporting of "bad" behaviors and states (e.g., smoking, drinking, stress) or over-reporting of "good" behaviors and states (e.g., healthy dieting, exercising, happiness).

Research on SD has been concerned with both its determinants as well as ways to deal with the bias to avoid that SD compromises the validity of study results. Determinants of SD are, among others, respondents' characteristics such as gender, age, and educational background (Crandall et al. 1965) as well as their personality traits, e.g., neuroticism (Brajsa-Zganec et al. 2011; Messina et al. 2010). With regards to age, past research has shown that the tendency to provide socially desirable answers is especially present in children, because when assessed by adults (e.g., by an adult interviewer) or in the presence of adults (e.g., at home or in the classroom), children tend to respond in a way that pleases adults to receive their social approval as they are physically, emotionally, and economically dependent and must maintain the good will of their caretakers (Crandall et al. 1965).

In addition, the magnitude of SD responding depends on the sensitivity of the researched behavior (Krumpal 2013), and issues with the data collection itself

(Nederhof 1985). Knowing what motivates SD responding allows to deal with this bias mainly in two ways: First, by employing different techniques during data collection, e.g. neutral questioning, randomized response technique, or self-administration, and, second, by assessing and controlling for SD bias with the help of scales developed and validated specifically for this purpose (Joinson 1999; Krumpal 2013; Nederhof 1985).

So far, only a few studies looked at the association between children's propensity to provide socially desirable answers, assessed with SD scales, and systematic over- or under-reporting about health-related issues. For example, Carifio found that "the tendency of middle school students to give socially desirable response is significantly correlated to their tendency to agree with myths about alcohol and drinking" (Carifio 1992, p.11). With regards to dietary reports, Guinn et al. (2010) could show that as SD scores increased, elementary school children tended to report less of the actually observed school-meal energy intake. Both studies used short versions of the Children's Social Desirability (CSD) scale (Crandall et al. 1965), which were derived through factor analyses.

The CSD is a 48-item scale for grades 6–12 children and a 46-item scale for grades 3–5 children with a binary response format, i.e., children either respond with "yes" or "no". Items cover both socially desirable (e.g., "Do you always listen to your parents?") and undesirable behaviors (e.g., "Have you ever felt like saying unkind things to a person?"). Carifio (1992) identified two parallel forms, each with twelve items from the original 48-item CSD. When administered to 181 sixth and eighth graders in the classroom, the two scales showed good internal consistency (Cronbach's alpha $\leq .70$) and 4-days test-retest reliability ($r \leq .83$). Guinn et al. (2010) used a 14-item subset of the 46-item CSD (Baxter et al. 2004). Psychometric analyses with data from 97 fourth grade children collected in the classroom showed good 1-month test-retest reliability ($r = .70$) and internal consistency at both time points (Cronbach alpha $\leq .82$) (Miller et al. 2014). Furthermore, a comparison of assessment modes among 157 fourth graders showed that both the interview (77 students) and the classroom (80 students) assessment produced good 1-month test-retest reliabilities ($r \leq .83$) and internal consistencies (Cronbach alpha $\leq .81$); the overall SD score was higher in the interview assessment (Miller et al. 2015).

As pointed out, the 14-item CSD-S was derived through factor analysis (FA). FA is commonly used to examine the dimensions of a multi-item scale and reduce scales to a set of items with maximal variability and reliability (Floyd and Widaman 1995). Though FA has become a popular technique in social sciences, it is limited to evaluate which items measure the same concept, i.e. share similar loadings or 'item difficulty'. FA does not take into account differences between the difficulty level of items for each person within a scale, i.e., differences in the probability that a person responds to one item compared to another (van Schuur 2003). Such differences are accounted for in Item Response Theory (IRT), which considers latent traits, e.g., SD, as a function of the persons' responses and the properties of the items that were administered (Embretson and Reise 2013). IRT has its roots in the development of psychological and cognitive abilities tests (Murphy and Davidshofer 2005; Rasch 1993). To our knowledge, IRT has not yet been applied to evaluate the properties of child-specific SD scales.

The present study has two main objectives: First, rather than challenging the short version of the CSD by Baxter et al. (2004), we take the 14-item scale and evaluate its unidimensionality and item properties applying IRT and thereby add valuable

information to the psychometric properties of the CSD-S. Second, we aim to contribute to research on the role of SD in self-report child well-being by looking at the dimension of social well-being, which, to the best of our knowledge, has not yet been related to SD responding when measured with the CSD-S.

As previously pointed out, short versions of the CSD correlated with children's self-report beliefs about alcohol myths and dietary reports, both linked to physical well-being. Another complex yet important dimension of well-being is social well-being. Individuals are "embedded in social structures and communities, and face countless social tasks and challenges" (Keyes 1998, p. 122). This already applies to children as they interact with other children and adults in their family and larger social environment. According to Keyes (1998), social well-being is comprised of social integration, social acceptance, social contribution, social actualization, and social coherence. For example, social integration describes a feeling of belonging to one's communities and society, whereas social acceptance encompasses feelings of trust and comfort, and a belief in the kindness of others. Social integration and social acceptance are incorporated in Putman's concept of social capital (Putnam et al. 1994), which consists of "trust, networks of cooperation and reciprocity, civic engagement and strong community identity" (Morrow 1999, p. 745).

Children's social capital is a resource that comes from within and outside the family (Coleman 1961). Accordingly, measures of social well-being in children assess, among others, family and peer relationships (Pollard and Lee 2003). Western societies value warm relationships with others and belonging. Because such culture-bound values are internalized from an early age (Grusec and Kuczynski 1997), it can be assumed that children in Switzerland high in SD report better family relationships.

H1: The higher children's tendency to give socially desirable responses, the more positive their reporting of family relationships.

As children grow and spend more time at school, peer relationships become increasingly important (Brown and Larson 2009). Peer relationships can be expressed by either behavior that is *performed* by children towards other children, e.g., friends, classmates, or *experienced* by children from others. Bad or deviant peer behavior can show in form of bullying, mobbing, or social isolation (Scholte and van Aken 2006). It can be expected that the frequency with which such bad behaviors are reported by children is negatively associated with SD.

H2: The higher children's tendency to give socially desirable responses, the less often they report on bad peer behaviors, either performed or experienced.

To sum up, the present paper aims to answer the following research questions:

*RQ1: What are the psychometric properties (*i.e.*, unidimensionality) of the CSD-S scale?*
*RQ2: What is the impact of SD, measured with the CSD-S, in addition to socio-demographic factors in children's self-report social well-being (*i.e.*, family and peer relationships)?*

2 Methods

2.1 Data Collection

Data come from the larger longitudinal *MEDIATICINO* study following students in their transition from childhood (9–10 years of age) to adolescence (14–15 years of age). SD bias was first assessed in the second wave. Data were collected between March and May 2015 among children attending fifth grade of public elementary school in Italian-speaking Switzerland. From 1184 sampled children, 1146 (97%) successfully completed the questionnaire. Given the territorial situation of Italian-speaking Switzerland, children came from small-city and village schools. In total, 66 grades distributed across 37 schools participated in wave 2. To increase the representativeness of the collected data, schools with higher student numbers collaborated with more classes, and students respectively. Teachers received a questionnaire for each student signed with a student identifier provided by the regional education administration, instructions for administration, and a pre-stamped return envelope. Teachers were asked to guide the self-administered completion of the questionnaire by reading aloud each question in class. Family background characteristics were available for a sub-sample of 843 students whose parents responded to a parent questionnaire signed with the same student identifier during the first wave between January and March 2014. An identifier was used to match the data. Since the use of an identifier assured anonymity of all data and thereby sufficiently addressed ethical considerations regarding privacy, the regional education administration approved the study design.

2.2 Measures

Social Desirability SD was assessed with the Children's Social Desirability Short (CSD-S) scale (Baxter et al. 2004; Miller et al. 2014, 2015). Two Italian mother-tongue translators with very good English skills translated the scale into Italian. Back-translation was conducted by two English mother-tongue translators with very good Italian skills to assure linguistic validity. The translated scale was furthermore pretested in five elementary school children for face validity. Items were formulated as questions to which students could respond either yes or no. Yes-answers were coded as 1 (socially desirable answer) and no-answers as 0 (socially undesirable answer). Before conducting data analyses, items assessing socially undesirable behavior were re-coded so that 1 indicated a socially desirable response across all items.

Family Relationships (Social Well-Being) Children's reported perceptions of positive family relationships were measured with six items according to Venkatraman et al. (2010), who applied them in a cross-cultural study including Italy. The Italian version of the items were obtained from the co-author who guided data collection in Italy. Each item (e.g., "The members of my family support each other.") was followed by a 4-point scale from 1 "completely disagree" to 4 "completely agree". All items were averaged to a compound score.

Peer Relationships (Social Well-Being) Children's social behavior with regards to peers was measured according to Mößle (2012). The German scale was translated into Italian. Back-translation by an independent translator and comparison with the original scale were conducted to assured linguistic validity. Eight items were formulated in active voice to measure *performed undesirable social behavior* (e.g., "I ignored another classmate on purpose."). The same eight items were re-formulated in passive voice to measure *experienced undesirable social behavior* (e.g., "Other classmates ignored me on purpose."). The scale ranged from 0 "never" to 3 "always" and measured the frequency with which the behavior was performed or experienced during the past month. Each group of items were averaged to created compound scores.

Socio-Demographic Characteristics Students' gender and age were assessed in the student questionnaire in wave 2. Males were coded as 0 and females as 1. Socio-economic background characteristics of their families were assessed in the parent questionnaire in wave 1. Parents were asked to indicate their annual gross household income in Swiss Francs on an ordinal scale with five increasing income classes. The highest educational attainment was assessed for each parent separately on an ordinal scale ranging from lower education to tertiary university education. The highest education attained among parents was used as an indicator of parental education level.

Univariate statistics of each final measure including mean, standard deviation, and Cronbach's alpha are reported in Table 3.

2.3 Data Analysis

Detailed psychometric analyses for the Italian Children's Social Desirability Short scale (I-CSD-S) were performed in R (R Core Team 2015) using the *mokken* package (Van der Ark 2007). More precisely, we examined the structural validity of the I-CSD-S using non-parametric Item Response Theory (NIRT), i.e. Mokken Scale Analysis (Mokken 1971; van Schuur 2003). A central assumption of MSA is that items are located on a continuum and have different endorsement frequencies (item difficulty). In case of the I-CSD-S, that means it is more likely that children give a socially desirable answer to one item than to another. According to MSA, items form a scale if they meet three conditions (e.g., Van der Ark 2012): (1) *unidimensionality*, i.e. the endorsement of a more difficult item should be related to an increased probability of endorsing less difficult items; (2) *local independence*, i.e. inter-item correlations should be explained only by the latent concept; and (3) *latent monotonicity* (the probability of endorsing an item should not decrease as scores on the latent concept increase. If these conditions are met, the item pool fits the monotone homogeneity model. If a fourth condition, i.e. *invariant item ordering* (items display the same order of difficulty irrespective of the scores on the latent concept) is met, the item set fits the double monotonicity model, which means all respondents perceive the item difficulties similarly and compound scale scores can be compared across subgroups (van Schuur 2003; Sijtsma and Hemker 1998). We tested the condition of unidimensionality by examining Loevinger's coefficients of homogeneity for each item of the I-CSD-S (H_i) and for the overall I-CSD-S (H). A minimum threshold of .30 should be achieved to retain items as part of the same latent concept and consider a scale as unidimensional (Ligtvoet et al. 2010). In addition,

an Automated Item Selection Procedure (aisp) algorithm available in the R *mokken* package was used to evaluate if all items of the I-CSD-S belong to a single latent concept at the minimum homogeneity threshold of .30. Monotonicity and invariant item ordering were tested via the check.monotonicity and check.restscore functions as part of the R *mokken* package. To evaluate if the binary items of the I-CSD-S increase monotonically and do not intersect, z-tests for each item should produce a significance value of 0 and a critical value below 40 (Sijtsma and Molenaar 2002).

In R, descriptive statistics were computed for all socio-demographics, the compound scores of social well-being indicators, and the compound score of the final I-CSD-S using the describe function in the R *psych* package (Revelle 2016). Reliability analysis was performed for all multi-item measures using the alpha function in the R *psych*, and a Cronbach's alpha of $\leq .70$ deemed to show good internal consistency (Nunnally 1978). Bivariate correlation analyses were performed via the corr.test function in the R *psych* package to test for the association between SD and socio-demographics as well as all three social well-being indicators.

Eventually, mixed-effects regression models were run using the R *lme4* package (Bates et al. 2015) to see if and, if so, how SD responding significantly biases children's self-report social well-being operationalized with perceived family and peer relationships where the term 'peer' refers to classmates. Mixed-effects models were used to account for the nested structure of the data that come from students sampled at grade level that are, in turn, nested in schools. Though the structure implies three-level data, two-level analyses were conducted as differences in social well-being, especially peer relationships, and SD responding (teacher-guided self-administration of the questionnaires in class) were expected at grade not school level.

3 Results

3.1 Sample

The final sample of this study consists of 843 students who took part in wave 1 and 2 of the longitudinal *MEDIATICINO* study and whose parents responded to a parent questionnaire in wave 1. This subset of students came from 55 grades distributed across 35 schools of different size. Schools with higher numbers of students were represented by two, in three cases three grades per school. As shown in Table 1, gender was distributed evenly. The modal highest educational attainment of parents was post-secondary non-tertiary education, i.e. most parents completed an apprenticeship, professional school, or high school. Slightly more than half of the families (56.4%) had an annual gross income of 96′000 Swiss Francs or less.

3.2 Psychometric Analyses of I-CSD-S

The number of missing values per item of the I-CSD-S ranged from 2 to 11 (see Table 2). Endorsement frequencies varied across all items. No item showed lack of variation, i.e. 10% or less of students providing a socially (un-)desirable answer (see Table 2). Thus, NIRT was conducted on a dataset with listwise excluded missing data (n = 799) for all 14 translated items from the original CSD-S scale.

Table 1 Sample characteristics (n = 843)

	n (%)	M (SD)
Child		
Male	413 (49.0)	
Female	430 (51.0)	
Age		10.4 (0.54)
Parents		
Highest educational attainment: at least one parent with…		
Lower secondary education	63 (7.5)	
Post-secondary non-tertiary education	444 (52.7)	
Tertiary applied university education	135 (16.0)	
Tertiary university education	188 (22.3)	
Not specified	13 (1.5)	
Gross annual household income in Swiss Francs		
Less than 48′000	92 (10.9)	
Between 48′000 and 72′000	205 (24.3)	
Between 72′001 and 96′000	179 (21.2)	
Between 96′001 and 120′000	141 (16.7)	
More than 120′000	145 (17.2)	
Not specified	81 (9.6)	

Loevinger's coefficient of homogeneity was below the minimum threshold of .30 for items I-CSD-S5 (H_i = .294), and I-CSD-S14 (H_i = .296). Examination of scale unidimensionality with increasing homogeneity thresholds via aisp indicated that item I-CSD-S5 did not belong to the latent concept at the minimum threshold of .30. After exclusion of this item, the coefficient of homogeneity for each remaining item was above .30 and for the overall 13-item scale .355 (SE = .016). The test of latent monotonicity revealed no violations since the z-tests for each item produced a significance value of 0 and a critical value below 40. The test of invariant item ordering via the check.restscore function in the R *mokken* package showed no intersecting item pairs. Thus, the 13-item scale also met all criteria for the double monotonicity model.

3.3 Correlation and Linear Mixed-Effects Regression Analyses

In R, on the full dataset including missing values (n = 843), a sum score was calculated and descriptive statistics run for the final I-CSD-S scale with a potential range from 0 to 13. The mean score was 5.32 (SD = 3.43). The scale was normally distributed and Cronbach's alpha of .82 showed good internal consistency (see Table 3). Missing values for the overall scale were present in 43 students (5.1%). The test of missingness at random (MAR) revealed that the presence of missing values in overall SD was significantly positively though weakly related to parent-report educational attainment, i.e. students who refused to answer to one or more items from the I-CSD-S, on average, came from families with higher educational

Table 2 Italian version of the children's social desirability short (I-CSD-S)

Item label	Italian version of CSD-S	Original version of CSD-S	SD response	# MV	Endorsement %
I-CSD-S1	Hai mai avuto voglia di dire cose cattive ad una persona?	Have you ever felt like saying unkind things to a person?	No	3	32.3
I-CSD-S2	Fai sempre attenzione a tenere i tuoi vestiti puliti e la tua camera ordinata?	Are you always careful about keeping your clothing neat and your room picked up?	Sì	10	59.8
I-CSD-S3	A volte preferisci stare a casa piuttosto che andare a scuola anche se non sei malato/a?	Do you sometimes feel like staying home from school even if you are not sick?	No	3	57.3
I-CSD-S4	Hai mai detto qualcosa che ha fatto sentire male un'altra persona?	Do you ever say anything that makes somebody else feel bad?	No	5	46.3
I-CSD-S5[a]	Sei sempre gentile, anche con le persone che non sono molto gentili con te?	Are you always polite, even to people who are not very nice?	Sì	10	49.6
I-CSD-S6	Qualche volta, fai delle cose anche se ti è stato detto di non farle?	Sometimes, do you do things you've been told not to do?	No	2	43.1
I-CSD-S7	Ascolti sempre i tuoi genitori?	Do you always listen to your parents?	Sì	5	54.4
I-CSD-S8	A volte desideri poter andare a giocare invece di andare a scuola?	Do you sometimes wish you could just play around instead of having to go to school?	No	5	37.6
I-CSD-S9	Hai mai infranto una regola?	Have you ever broken a rule?	No	6	39.4
I-CSD-S10	Qualche volta ti arrabbi quando non puoi fare a modo tuo?	Do you sometimes feel angry when you don't get your way?	No	7	24.7
I-CSD-S11	Qualche volta prendi in giro altre persone?	Do you sometimes feel like making fun of other people?	No	7	47.1
I-CSD-S12	Fai sempre la cosa giusta?	Do you always do the right things?	Sì	11	22.7
I-CSD-S13	Ci sono delle volte che non ti piace fare quello che ti dicono i tuoi genitori?	Are there some times when you don't like to do what your parents tell you?	No	5	14.9
I-CSD-S14	Qualche volta ti arrabbi quando le persone non fanno quello che vuoi tu?	Do you sometimes get mad when people don't do what you want them to do?	No	7	52.9

$n = 843$; [a] item excluded from final I-CSD-S scale

background. The lack of MAR was accounted for as parental educational attainment was in any case included as a covariate in the mixed-effects regression analyses (Acock 2005). Given the considerable amount of missing data, especially for family income (see Table 3), a multiple imputation procedure as suggested by Myers (2011) was chosen. Using the *mitml* package in R (Grund et al. 2016), 100 imputed data sets provided the basis for all linear mixed-effects regression models described hereafter.

Bivariate correlation analyses revealed highly significant correlations between SD and child report family relationships ($r = .302$, $p < .001$) as well as peer relationships, i.e. performed undesirable behavior towards other children ($r = -.504$, $p < .001$) and

Table 3 Descriptive statistics for manifest indicators and bivariate correlations with I-CSD-S

	# items	Range	# MV	M	SD	Skew	Kurt	Alpha (95% CI)	I-CSD-S	
									r	*p*
I-CSD-S	13	0–13	43	5.32	3.43	0.33	−0.82	0.82	-	-
Gender (Female)	1	0/1	0					-	.131	< *.001*
Parental education	1	1–4	13	2.54	0.92	0.46	−0.93	-	.026	*.463*
Family income	1	1–5	81	3.06	1.30	0.09	−1.14	-	.054	*.148*
Social well-being										
Family relationships	6	1.17–4	3	3.49	0.51	−1.04	0.72	0.80	.302	< *.001*
Peer relationships: performed undesirable behavior	8	0–1.62	14	0.18	0.21	1.84	5.50	0.64	−.504	< *.001*
Peer relationships: experienced undesirable behavior	8	0–3	12	0.21	0.33	3.39	18.24	0.79	−.131	< *.001*

n = 843; Pearson's r based on Holm correction for multiple tests

experienced undesirable behavior from other children ($r = -.131$, $p < .001$). The direction of the coefficients was as hypothesized, i.e., "good" indicators correlated positively and "bad" indicators negatively with SD (Table 3). Table 3 also shows the coefficients for the correlations of the I-CSD-S with socio-demographic variables. SD was not related to children's socio-economic background, but to gender ($r = .131$; $p < .001$). Girls ($M = 5.76$, $SD = 3.42$), on average, provided more socially desirable answers than boys ($M = 4.86$, $SD = 3.39$) ($t(798) = -3.74$, $p < .001$).

Linear mixed-effects regression models were run for each indicator of social well-being to determine the impact of SD bias on these indicators in addition to socio-demographic child characteristics. The models accounted for the nested structure of the data, and allowed for variation in social well-being at grade level. With regards to family relationships, children's tendency to provide socially desirable answers significantly positively predicted their report of trust, belonging, and support within the family ($B = 0.04$, $t(49272) = 7.45$, $p < .001$), more and above children's socio-demographic characteristics (*H1* confirmed). Furthermore, controlling for gender and a family's educational and financial resources, SD significantly negatively predicted children's self-report undesirable behavior towards their classmates ($B = -0.03$, $t(39927) = -14.40$, $p < .001$) and experienced undesirable behavior from other classmates ($B = -.0.01$, $t(39028) = -2.86$, $p = .002$) (*H2* confirmed) (Table 4). Children's propensity to SD responding explained additional 7% of variance in family relationships, 1% in experienced undesirable peer behavior, and 22% in undesirable peer behavior performed by the respondents themselves.

Models 2 in Table 4 also include a fixed effect of SD bias at grade level to test for potential effects of the average grade SD score on children's social well-being as one could expect that the presence of the teacher during self-administration may impact students' responses and, thus, contribute to differences in the reporting of social well-being between grades (between-group differences). However, the non-significant effects exclude such a consideration and underline that SD biases self-report perceptions of family and peer relationships at the individual level.

Table 4 Results of linear mixed-effects regression models predicting child-report social well-being

	Family relationships						Peer relationships (Performed undesirable behavior)						Peer relationships (Experienced undesirable behavior)					
	Model 1			Model 2			Model 1			Model 2			Model 1			Model 2		
Fixed	*B*	*B SE*	*p*	*B*	*B SE*	*p*	*B*	*B SE*	*p*	*B*	*B SE*	*p*	*B*	*B SE*	*p*	*B*	*B SE*	*p*
Intercept	3.31	0.06	*<.001*	3.09	0.09	*<.001*	0.25	0.03	*<.001*	0.39	0.03	*<.001*	0.31	0.04	*<.001*	0.38	0.05	*<.001*
Socio-demographics																		
Gender (Female)	0.10	0.04	*.001*	0.07	0.03	*.025*	−0.08	0.01	*<.001*	−0.05	0.01	*<.001*	−0.05	0.02	*.016*	−0.04	0.02	*.046*
Highest parental education	−0.01	0.02	*.307*	−0.01	0.02	*.299*	−0.01	0.01	*.071*	−0.01	0.01	*.043*	−0.03	0.01	*.012*	−0.03	0.01	*.011*
Annual family household income	0.05	0.02	*.001*	0.05	0.02	*.001*	0.00	0.01	*.333*	0.00	0.01	*.222*	0.00	0.01	*.392*	0.01	0.01	*.324*
Social desirability (individual level)				0.04	0.01	*<.001*				−0.03	0.00	*<.001*				−0.01	0.00	*.002*
Social desirability (grade level)				0.00	0.01	*.405*				0.00	0.01	*.408*				0.00	0.01	*.288*
Δ Adjusted R^2 (individual level)				.07						.22						.01		
Random	*B*			*B*			*B*			*B*			*B*			*B*		
Class	0.015			0.010			0.003			0.001			0.001			0.001		
Student	0.243			0.226			0.041			0.032			0.104			0.103		
ICC	0.056			0.042			0.074			0.041			0.014			0.011		
F value	33.54						105.96						6.54					
p	*< .001*						*< .001*						*.001*					

n = 843; results based on multiple imputations of multilevel missing data

4 Discussion

Research on child well-being largely relies on self-report data either due to the nature of the well-being indicator, i.e., when the indicator can only be assessed through self-report, or due to financial and time resources, which require researchers to rely on self-report even when non-reactive methods are at hand. A shortcoming of self-report data is the potential systematic error, called "bias", which is introduced by children's tendency to provide socially desirable answers when asked about sensitive topics or behaviors, e.g., substance use or deviant social behavior (Ford and Rubin 1970). One way to deal with social desirability (SD) is to measure and control for the bias in subsequent statistical analyses (Nederhof 1985). To date, only a few child-specific SD scales have been developed and validated, including the 14-item Children's Social Desirability Short (CSD-S) scale (Baxter et al. 2004), which was derived from the 46-item binary scale by Crandall et al. (1965). In the present study, we analyzed the psychometric properties of an Italian version of the CSD-S by using nonparametric Item Response Theory (NIRT) and reliability analysis. Based on data from fifth grade elementary school children in Italian-speaking Switzerland (aged 10), we identified a 13-item I-CSD-S scale that met the double monotonicity model, i.e. all items belonged to one scale (unidimensionality), and each item increased monotonically and did not intersect with other items from the scale. Given that this is the first study applying NIRT to examine the psychometric properties of the CSD-S, comparison of the present results confirming double monotonicity with the original 14-item scale is not possible. However, internal consistency of both scales could be compared and showed that the 13-item I-CSD-S produced an equally satisfactory Cronbach's alpha (> .80).

In contrast to the 13-item I-CSD-S, children's reported peer relationships in the present study were not normally distributed, which is indicative of some form of bias in self-report social well-being. In fact, when controlling for gender and socio-economic background, SD was significantly associated with self-report indicators of family and peer relationships. As hypothesized, indicators with a positive connotation were positively associated with SD (i.e., perceived family relationships) and those with a negative connotation negatively (i.e., undesirable performed and experience peer behaviors). This confirms that children tend to over-report "good" and under-report "bad" behaviors or states (Ford and Rubin 1970). Though SD explained only little additional variance in self-report family relationships and experienced undesirable behaviors by peers, the significant associations underline the need to account for SD in self-report studies on child well-being. Especially with regards to self-report performed undesirable behavior towards peers, where SD explained additional 22%, the bias should be taken into account. Also, the effect of individual-level SD bias in self-report experienced peer behavior is three times smaller than the effect in performed behavior. That means, children are considerably more willing to report on how much bullying actually happens to them compared to how much they engage in it, though past research has identified several reasons why victims of deviant peer behavior hesitate to disclose (Mishna and Alaggia 2005).

The findings of this study call for child-specific SD scales in national and international surveys like the Health Behavior in School-aged Children (HBSC) (www.hbsc.org) or the International Survey of Children's Well-Being (ISCWeB) (www.isciweb.org), which are frequently used to conduct multivariate and comparative secondary data

analyses. It should be noticed, though, that the assessment and control of SD bias in statistical analyses do not account for possible differences in children's subjective perception of personal well-being. That is, children may vary in their responses to (social) well-being indicators also because they experience them differently or assign to them different weights. Since a common cause for such a "subjectivity bias" is hard if not impossible to identify, this systematic bias remains inherent to self-report data.

In addition to the main findings, the present study also revealed that the tendency to give socially desirable answers is higher in girls than in boys. This is in line with results from previous studies using the original CSD, and it confirms that girls are more in search for social approval when asked about sensitive topics (Crandall et al. 1965; Klein et al. 1969). Given the relationship between gender and SD, when controlling for the latter, girls significantly report better family relationships and less undesirable peer relationships than their male counterparts. These results highlight that boys are more at risk for negative consequences in their development and social well-being (Gove 1985; Young and Sweeting 2004), and the gender differences are not simply an artifact of SD trait.

Eventually, the multilevel results of this study deserve further discussion. The non-significant effect of SD bias at class level underlines that SD biases self-report perceptions of family and peer relationships at the individual level, independently from students' affiliation to a specific learning community and their relationship to the teacher in class. This finding refutes concerns regarding systematic bias due to the presence of teachers during data collection. In fact, self-administration in classroom settings is a frequent method of data collection among children as schools provide a structured and easy to control access to children populations. Furthermore, self-administration in class is less time and cost consuming and produces lower SD biased responses than personal interviews (Miller et al. 2015). To further reduce potential SD bias, researchers are advised to carefully design the questionnaire (Krumpal 2013), paying special attention to the introductory explanations. These should clearly state that the questionnaire is not a test and that all responses are treated confidentially meaning that neither parents nor teachers will have access to the data. Such statements are usually anyway required for ethical approval.

To conclude, some limitations of the present study should be acknowledged. First, the broader scope of the longitudinal study from which the data come from does not allow a more detailed assessment of social well-being its association with children's socio-demographic characteristics and tendency to SD responding. For example, Young and Sweeting (2004) point out that deviant peer behavior can have different forms for boys and girls. While boys engage more in physical forms of bullying, girls use more verbal harassment. By differentiating between these two forms, it would be interesting to see if children's self-report is equally impacted by SD bias, controlling for gender, or not. Second, the present study did not examine the role of SD bias in an extended conceptualization of child well-being by including dimensions other than social well-being and associated sensitive multi-item indicators such as smoking or drinking. The collaborating regional education administration did not approve the assessment of substance use in the young age group (aged 10) of the present sample. However, data from the 2014 HBSC survey in Switzerland showed that among 11-year olds, 20.1% of the boys and 10.5% of the girls have already drunk alcohol (Marmet et al. 2015). These numbers call for a thorough evaluation of SD bias in self-report

measures on substance use already at a younger age. Third and last, the data for this study is not directly comparable to past findings on the original CSD-S scale, which was developed and validated in an American context and linked to physical well-being indicators (e.g., dietary reports), whereas the present study examined an Italian version of the CSD-S in a Swiss context and associated children's scores on the scale to social well-being indicators. Though this was not the focus of the present study, a comparative cross-country study would provide valuable insights in the role of cultural context in SD bias and self-report child well-being. In their study on values, Fisher and Katz (2000) showed that the magnitude of SD bias in reported values indicates the culture-bound importance of individual values. These values include, among others, warm relationships with others and sense of belonging. When applied to the personal situation, both represent indicators of social well-being. Thus, more research on SD and child indicators is needed.

Acknowledgements We wish to thank the regional education administration of Ticino and the participating schools for their collaboration during data collection. We also thank Dr. A.L. Dima for her constructive feedback on Item Response Theory analysis.

Compliance with Ethical Standards

Conflict of Interest None declared.

Funding Internal funding.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family, 67*(4), 1012–1028. doi:10.1111/j.1741-3737.2005.00191.x.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48 arXiv:1406.5823.

Baxter, S. D., Smith, A. F., Litaker, M. S., Baglio, M. L., Guinn, C. H., & Shaffer, N. M. (2004). Children's social desirability and dietary reports. *Journal of Nutrition Education and Behavior, 36*(2), 84–89. doi:10.1016/S1499-4046(06)60138-3.

Bradshaw, J., Hoelscher, P., & Richardson, D. (2006). An index of child well-being in the European Union. *Social Indicators Research, 80*(1), 133–177. doi:10.1007/s11205-006-9024-z.

Brajsa-Zganec, A., Ivanovic, D., & Lipovcan, L. K. (2011). Personality traits and social desirability as predictors of subjective well-being. *Psychological Topics, 20*(2), 261–276.

Brown, B. B., & Larson, J. (2009). Peer relationships in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: volume 2 contextual influences on adolescent development* (3rd ed., pp. 74–103). Hoboken: John Wiley & Sons, Inc.

Carifio, J. (1992). Parallel short forms of the Crandall social desirability test for children: shortening instruments for research purposes. *Paper Presented at the Annual Conference of the Eastern Educational Research Association, Hilton Head, So. Carolina, March 5–9, 1992*. Retrieved from http://eric.ed.gov.ezp.lib.rochester.edu/?id=ED343915.

Coleman, J. S. (1961). *The adolescent society: the social life of the teenager and its impact on education*. New York: Free Press on Glencoe.

Crandall, V. C., Crandall, V. J., & Katkovsky, W. (1965). A children's social desirability questionnaire. *Journal of Consulting Psychology, 29*(1), 27–36. doi:10.1037/h0020966.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. doi:10.1037/h0047358.

Earls, F., & Carlson, M. (2001). The social ecology of child health and well-being. *Annual Review of Public Health, 22*(1), 143–166. doi:10.1146/annurev.publhealth.22.1.143.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden Press.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Hove: Psychology Press.

Fisher, R., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing, 17*(2), 105–120. doi:10.1002/(SICI)1520-6793(200002)17:2<105::AID-MAR3>3.0.CO;2-9.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286–299. doi:10.1037/1040-3590.7.3.286.

Ford, L. H., & Rubin, B. M. (1970). A social desirability questionnaire for young children. *Journal of Consulting and Clinical Psychology, 35*(2), 195–204. doi:10.1037/h0030042.

Gove, W. R. (1985). The effect of age and gender on deviant behavior: a biopsychosocial perspective. In A. S. Rossi (Ed.), *Gender and the life course* (pp. 115–144). Hawthorne: Aldine de Gruyter.

Grund, S., Lüdtke, O., & Robitzsch, A. (2016). Multiple imputation of multilevel missing data: an introduction to the R package pan. *SAGE Open, 6*(4), 1–17. doi:10.1177/2158244016668220.

Grusec, J. E., & Kuczynski, L. E. (1997). Parenting and children's internalization of values: a handbook of contemporary theory. Hoboken NJ: John Wiley & Sons Inc.

Guinn, C. H., Baxter, S. D., Royer, J. A., Hardin, J. W., Mackelprang, A. J., & Smith, A. F. (2010). Fourth-grade children's dietary recall accuracy for energy intake at school meals differs by social desirability and body mass index percentile in a study concerning retention interval. *Journal of Health Psychology, 15*(4), 505–514. doi:10.1177/1359105309353814.

Hauser, R. M., Brown, B. V., & Prosser, W. R. (1997). *Indicators of children's well-being*. New York: Russell Sage Foundation.

Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*(3), 433–438. doi:10.3758/BF03200723.

Keyes, C. L. M. (1998). Social well-being. *Social Psychology Quarterly, 61*(2), 121–140. doi:10.2307/2787065.

Klein, E. B., Gould, L. J., & Corey, M. (1969). Social desirability in children: an extension and replication. *Journal of Consulting and Clinical Psychology, 33*(1), 128–128. doi:10.1037/h0027339.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity, 47*(4), 2025–2047. doi:10.1007/s11135-011-9640-9.

Ligtvoet, R., van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement, 70*(4), 578–595. doi:10.1177/0013164409355697.

Marlowe, D., & Crowne, D. P. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology, 25*(2), 109–115. doi:10.1037/h0041627.

Marmet, S., Archimi, A., Windlin, B., & Delgrande Jordan, M. (2015). *Substance use among pupils in Switzerland in 2014 and trends since 1986 – results of the "health behaviour in school-aged children" (HBSC) study*. Lausanne: Addiction Switzerland Retrieved from http://www.addictionsuisse.ch/fileadmin/user_upload/DocUpload/Summary-HBSC-2014.pdf.

Messina, A., Fogliani, A. M., & Paradiso, S. (2010). Association between alexithymia, neuroticism, and social desirability scores among Italian graduate students. *Psychological Reports, 107*(1), 185–192. doi:10.2466/02.07.13.15.PR0.107.4.185-192.

Miller, P. H., Baxter, S. D., Hitchcock, D. B., Royer, J. A., Smith, A. F., & Guinn, C. H. (2014). Test-retest reliability of a short form of the children's social desirability scale for nutrition and health-related research. *Journal of Nutrition Education and Behavior, 46*(5), 423–428. doi:10.1016/j.jneb.2013.11.002.

Miller, P. H., Baxter, S. D., Royer, J. A., Hitchcock, D. B., Smith, A. F., Collins, K. L., & Finney, C. J. (2015). Children's social desirability: effects of test assessment mode. *Personality and Individual Differences, 83*, 85–90. doi:10.1016/j.paid.2015.03.039.

Mishna, F., & Alaggia, R. (2005). Weighing the risks: a child's decision to disclose peer victimization. *Children & Schools, 27*(4), 217–226. doi:10.1093/cs/27.4.217.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: with applications in political research*. Berlin: Walter de Gruyter.

Morrow, V. (1999). Conceptualising social capital in relation to the well-being of children and young people: a critical review. *The Sociological Review, 47*(4), 744–765. doi:10.1111/1467-954X.00194.

Mößle, T. (2012). "*Dick, dumm, abhängig, gewalttätig?*" *Problematische Mediennutzungsmuster und ihre Folgen im Kindesalter. Ergebnisse des Berliner Längsschnitt Medien (Vol. 42)*. Baden-Baden: Nomos.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing. Principles and applications* (6th ed.). New Jersey: Prentice Hall.

Myers, T. A. (2011). Goodbye, listwise deletion: presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures, 5*(4), 297–310. doi:10.1080/19312458.2011.624490.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: a review. *European Journal of Social Psychology, 15*(3), 263–280. doi:10.1002/ejsp.2420150303.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*(1), 1–18. doi:10.1016/0022-2496(66)90002-2.

Nunnally, J. (1978). *Psychometric methods*. New York: McGraw-Hill.

Pollard, E. L., & Lee, P. D. (2003). Child well-being: a systematic review of the literature. *Social Indicators Research, 61*(1), 59–78. doi:10.1023/A:1021284215801.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: civic traditions in modern Italy*. Princeton: Princeton University Press.

R Core Team. (2015). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press.

Revelle, W. (2016). *Psych: procedures for personality and psychological research*. Evanston: Northwestern University. https://CRAN.R-project.org/package=psych , Version= 1.5. 8.

Scholte, R. H. J., & van Aken, M. (2006). Peer relations in adolescence. In S. Jackson & L. Goossens (Eds.), *Handbook of adolescent development* (pp. 175–199). New York: Psychology Press.

Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika, 63*(2), 183–200. doi:10.1007/BF02294774.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (5th ed.). Thousand Oaks: Sage.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 20*(11), 1–19. doi:10.18637/jss.v020.i11.

Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software, 48*(5), 1–27. doi:10.18637/jss.v048.i05.

van Schuur, W. H. (2003). Mokken scale analysis: between the Guttman scale and parametric item response theory. *Political Analysis, 11*(2), 139–163. doi:10.1093/pan/mpg002.

Venkatraman, S., Dishion, T. J., Kiesner, J., & Poulin, F. (2010). Cross-cultural analysis of parental monitoring and adolescent problem behavior: theoretical challenges and model replication when east meets west. In V. Guilamo-Ramos, J. Jaccard, & P. Dittus (Eds.), *Parental monitoring of adolescents: current perspectives for researchers and practitioners*. New York: Columbia University Press.

Young, R., & Sweeting, H. (2004). Adolescent bullying, relationships, psychological well-being, and gender-atypical behavior: a gender diagnosticity approach. *Sex Roles, 50*(7–8), 525–537. doi:10.1023/B:SERS.0000023072.53886.86.