Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 02-2022

Marina Lagemann and Peter Winker

Inconsistent Response Behavior: A Potential Pitfall in Modeling the Link between Educational Attainment and Social Network Characteristics

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Inconsistent response behavior: A potential pitfall in modeling the link between educational attainment and social network characteristics[1]

Marina Lagemann[2] and Peter Winker[3]

Abstract

An important role is ascribed to students' social networks in explaining both social and ethnic differentials in educational achievement and attainment. For example, students' social networks are assumed to influence their probability of success by providing educationally-relevant resources and by promoting effort and educational investments. The direction and strength of the network's effect on students' educational success is assumed to depend on the network's precise characteristics, such as educational and migration background. As track selection by school performance (as is the case in Germany) goes hand in hand with a segregation of students by characteristics like social and migration background, it can be assumed that educational success itself has an influence on the social resources students have access to at later stages of their educational careers. Given the complexity of instruments commonly applied in self-administered questionnaires to assess students' social resources, the quality of data on measures of network characteristics is likely to depend on the respondents' abilities. As regards the estimation of the association between network characteristics and educational success, biased measurement of social network characteristics apparently constitutes a challenge as spurious correlation may be observed between measures of educational achievement and network characteristics if the bias systematically correlates with education.

We report empirical findings on a complex instrument used in a self-administered questionnaire applied in the National Educational Panel Study (NEPS) to 9th-graders in the classroom, which was designed to measure the social resources young people have at their disposal at the point of transition from general into vocational education. The data allows identifying population subgroups who face particularly strong difficulties in completing the relevant set of questions in a consistent way. Specifically, this selection can be shown to be significantly correlated with different measures of educational achievement as well as with the respondents' migration background. As the network characteristics we investigate, i.e., the network members' educational and migration background, have been found to correlate with students' educational success, ignoring this selection can be shown to heavily bias estimates of the association between educational achievement and social network characteristics.

Keywords: Social networks; network characteristics; network composition; social resources; answering behavior; cognitive skills; measurement bias; migration background; educational success; educational attainment

---

1 This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 9, doi:10.5157/NEPS:SC4:7.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.
The German Federal Ministry of Education and Research provided funding for the research project REDMig – "Non-monetary returns to education in the form of social inclusion: Estimation and interdependence of private and social returns to education of migrants and non-migrants" (funding ID: RDMig2, Prof. Dr. Peter Winker).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
2 University of Giessen, Faculty of Social Sciences and Cultural Studies, Karl-Glöckner-Str. 21B, Germany, email: marina.lagemann@erziehung.uni-giessen.de.
3 University of Giessen, Faculty of Economics and Business Studies, Licher Strasse 64, Germany, email: Peter.Winker@wirtschaft.uni-giessen.de, corresponding author.

1. Introduction

The analysis of social networks has become an issue of high relevance in different fields of social sciences.[4] Given the presumed role of students' social networks in influencing educational success, the analysis of the social resources students have at their disposal at the different points of transition in the educational system as well as of their effects and determinants is of central interest for the analysis of educational inequalities. For instance, students' social networks are assumed to influence educational outcomes by means of providing educationally-relevant resources, by promoting motivation and effort and by influencing educational decisions (Becker & Gresch 2016).

It is assumed that the direction and strength of the network's effect on students' educational success depends on the network's precise characteristics. Examples are the network members' educational background, migration background and educational motivation (Bygren und Szulkin 2010; Kroneberg 2008): Higher educated networks will be more able to provide support (e.g., financially and in terms of providing help with schoolwork and relevant information) and more likely to hold high aspirations and expectations (Manski 2004; Morgan 1998, 2002). The influence of the network's migration background is discussed controversially. On the one hand, migrants may be less able to provide educationally-relevant support than their native counterparts (even with a comparable educational background) due to migration-specific circumstances such as language barriers and lacking experience with the educational system in the receiving country. On the other hand, the comparatively high educational aspirations and expectations of many migrant groups may positively influence students' investment behavior (Becker & Gresch 2016).

Also, it can be assumed that it is not only the case that students' social resources influence students' educational success, but also that educational success itself influences the social resources students have access to in the future. For instance, the selection of students into tracks by school performance goes hand in hand with a segregation by other characteristics like social, educational and migration background (Baumert, Stanat, Watermann 2006).[5] As such, prior educational performance and decisions are likely to have an effect on the social resources students can access at later stages of their educational careers: Students in more prestigious tracks will be more likely to make friends with classmates from higher social and educational backgrounds and without a migration background (in particular migrants who are underrepresented in higher tracks, like students of Turkish origin in Germany).

Different approaches have been developed for operationalizing and collecting egocentric social network data in a survey setting that take into account both the respondents' burden and the time constraints for interviews. A common approach of larger scale studies is to collect information about an a priori limited number of network members (e.g., NEPS, SOEP).[6] Especially when using complex instruments to assess social resources, measures from self-administered questionnaires may be biased by the respondents' cognitive capacities (Bradburn 1979, Lenzener et al. 2010). Further, it can be assumed that migrants have more difficulty with consistently filling out questionnaires due to migration-specific conditions such as lower majority language skills (Becker 2010). As regards the estimation of the association between network characteristics and educational success, biased measurement of social network characteristics apparently constitutes a challenge as spurious

[4] For a broad coverage see, e.g., Scott and Carrington (2014).

[5] This does not only apply to the selection into different school forms (like in Germany) but also to educational systems without track selection when students are selected or select themselves into courses of different prestige within one and the same school (Heck, Price, & Thomas, 2004; Oakes & Wells, 1996; Oakes u.a., 1990; Baumert, Stanat, Watermann 2006).

[6] For a detailed description of the National Educational Panel Study (NEPS) for Germany see Fuß et al. (2016). Details about the German Socio-Economic Panel Study (SOEP) are provided by Wagner et al. (2007).

correlation may be observed between measures of educational achievement and network characteristics if the bias systematically correlates with education.

In our study, we specifically investigate the response behavior of secondary students participating in the National Educational Panel Study (NEPS), which provides longitudinal data on educational processes and competence development in Germany and, as such, constitutes a major source for the investigation of the (re)production of educational inequalities. We focus on the investigation of inconsistencies that arise when questions on the social resources students have at their disposal to successfully manage the transition from general into vocational education. Specifically, we address the question whether inconsistent response behavior in terms of not filling out the relevant sets of questions correctly systematically depends on students' educational achievement and migration background, which have been identified as network characteristics that influence the amount and quality of support students receive in the course of their educational careers (Roth, Salikutluk and Kogan 2010). Our first objective is to investigate inconsistent answering behavior by education, specifically track attendance, and migration background. Our second objective is to investigate the factors that drive these inconsistencies and – by means of explicitly considering these factors in our estimations – to illustrate how (ignoring) response errors made by the participants when filling out questionnaires may systematically bias the estimated association between educational achievement and network characteristics.

The paper is structured as follows: Section 2 introduces the NEPS data, focusing on the questions asked to elicit information on students' social resources at the transition from general into vocational education. The participants' response behavior is presented and discussed with respect to observed inconsistencies in Section 3. To this end, we provide descriptive statistics for different subgroups of respondents depending on whether or not they managed to complete the relevant questions on network size and structure in a consistent way and specify and estimate an empirical model for the observed answering behavior. In Section 4, we empirically demonstrate how inference depends on whether the selection bias is ignored or explicitly modeled when estimating the link between education and network characteristics. We discuss our results in Section 5.

2. Data and methodological approach

Our analyses are based on data from the National Educational Panel Study (NEPS),[7] a longitudinal study that follows six cohorts of over 60,000 persons (starting in 2009) in Germany through the different stages of their educational careers from new-borns to adults. The study aims at providing detailed insight into the development of competencies, educational decisions and processes as well as into the returns to education across the lifespan. Besides questionnaire data on aspects like determinants of and returns to education (such as social networks, learning environments and labor market outcomes), the NEPS provides competence data in different domains (e.g., von Maurice, Leopold & Blossfeld 2011; von Maurice, Sixt & Blossfeld 2011).

The present study uses both questionnaire and competence data from the first and second wave of the fourth starting cohort, which comprises 9th-graders in the first and second half of the school year. The first wave was carried out from November 2010 to January 2011, the second wave from May to July 2011. The questionnaire and test data used in our analyses was collected in school by means of

[7] The NEPS is funded by the Federal Ministry of Education and Research and carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg.

self-administered questionnaires by the IEA DPC – IEA Data Processing and Research Center, Hamburg.[8]

The data we analyze in detail are students' responses to two consecutive sets of questions with identical structure that were designed to assess the social resources students have at their disposal at the point of transition into vocational education and training (VET) in terms of knowing people that (1) can [or would potentially] provide information about available positions in VET on the one hand (first set of questions, referred to as *info jobs* from here on), and (2) make actual effort to help the target person to actually get a position in VET on the other hand (second set of questions, referred to as *help jobs* from here on).

The first set of questions is introduced as follows: "The following questions are about the people in your life, whether you know them well or not. Imagine you're looking for a vocational training position. How likely is it that people in your life would let you know about interesting open vocational training positions?" The second set of questions starts off with "Imagine you're looking for a vocational training position. How likely is it that someone in your life would help you write an application for that vocational training position?"[9] Following this introduction, students were asked to report the probability to know at least one person who would provide information about interesting VET positions (*info jobs*) and who would make effort to get a VET position for them (*help jobs*) based on a 4-point Likert Scale ranging from "very unlikely" to "very likely". Students who reported to be "very" or "rather likely" to know at least one such person were further asked to indicate the specific person(s) (e.g., mother, sibling) and the network size, i.e., total number of persons they had in mind when answering the preceding two questions (answer categories "one", "two", "three or more"). The probability to know somebody and the network size are assessed separately for *info jobs* and *help jobs*; see question number 22, Figure 1). Finally, the participants were asked to provide information on different network characteristics. In our study we focus the network share who has (a) completed the highest level of general education (*Abitur*) and (b) a migration background (referring to persons who themselves and/or of whom at least one parent was born abroad) (see question number 22, Figure 1).

For our analyses, we use the operationalization of these shares as provided by the NEPS in order to be able to compare network characteristics across networks of different size: If the respondent reports to know one person, the possible outcomes are "yes" (i.e., person has a migration background/obtained the *Abitur*, coded as 7) and "no" (coded as 1). For two persons, the possible outcomes are "none" (coded as 1), "one" (coded as 4) and "both" (coded as 7). For three or more persons, there are seven possible outcomes: "none" (coded as 1), "almost none" (coded as 2), "less than half" (coded as 3), "approximately half" (coded as 4), "over half" (coded as 5), "almost all" (coded as 6) and "all" (coded as 7).

---

[8] A full documentation of the data collection including the instruments used can be found at https://www.neps-data.de/tabid/845/language/en-US/Default.aspx. Note that the original layout of the questionnaire as used in school is provided in German only. The German questionnaire (field version) can be downloaded at https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/Feldversionen/SC4_Q_w1-2_de.pdf. The English translation (SUF Version) can be downloaded at https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/1-1-0/SC4_1-1-0_Q_w1_2_en.pdf.

[9] The original German wording of the two questions in the field versions are: "Stell dir vor, du suchst einen Ausbildungsplatz. Wie wahrscheinlich ist es, dass dich jemand in deinem persönlichen Umfeld über interessante freie Ausbildungsplätze informieren würde" (question 20) and „Wie wahrscheinlich ist es, dass sich jemand aus deinem persönlichen Umfeld dafür einsetzen würde, dass du einen Ausbildungsplatz bekommst?" (question 22).

Figure 1: Excerpt from NEPS questionnaire.

22 All in all, how many people did you have in mind when you heard the last two questions?		
Please check only one answer		
one person ☐ ↓ *Please answer the questions in this column only.*	two people ☐ ↓ *Please answer the questions in this column only.*	more than two people ☐ ↓ *Please answer the questions in this column only.*
Does this person have an immigration background?	**How many of them have an immigration background?**	**How many of them have an immigration background?**
Immigration background means that the person themselves or at least one parent of that person was born abroad.		
Please check where applicable.	*Please check only one answer.*	*Please check only one answer.*
yes ☐	both ☐	none of them ☐
no ☐	one ☐	almost none of them ☐
	none ☐	less than half of them ☐
		approximately half of them ☐
		over half of them ☐
		almost all of them ☐
		all of them ☐
And does this person have the Abitur?	**And how many of them have the Abitur?**	**And how many of them have the Abitur?**
Please check where applicable.	*Please check only one answer.*	*Please check only one answer.*
yes ☐	both ☐	none of them ☐
no ☐	one ☐	almost none of them ☐
	none ☐	less than half of them ☐
		approximately half of them ☐
		over half of them ☐
		almost all of them ☐
		all of them ☐

As measures for level of educational achievement of the target persons, we use information on the type of school attended as well as competence data. In the German educational system, after four years of primary education students are traditionally selected into three different tracks of secondary education that lead to distinct school leaving certificates, which in turn provide access to different types of vocational and post-secondary education: *Hauptschule* ('lower secondary education') traditionally leads to the lowest qualification after nine years of general education, *Realschule* ('intermediate secondary education') leads to a medium-level education after ten years of general education, and *Gymnasium* ('higher secondary education') leads to a qualification that provides access to higher education after 12 or 13 years of general education.

Apart from the most demanding track, which in fact constitutes the only constant in the educational system across the federal states (*Länder*), this traditional division has been progressively replaced over the last years by schools that offer several or all tracks of secondary education (Trebbels 2014; cf. also Stubbe et al. 2009, Stubbe et al. 2012). The variable we are using in our analyses has the outcomes "Hauptschule", "Realschule", "Gymnasium" as well as two types of schools that offer several tracks of secondary education: "Integrierte Gesamtschule" (*Integrated School*) offers all tracks, "Schule mit mehreren Bildungsgängen" (*Comprehensive School*) offers all tracks except for the academic track that leads to the highest general educational qualification (*Abitur*) (cf. Skopek, Pink & Bela 2013).

Further, we use the results from different competence tests that were administered directly on the students. Specifically, we use measures of students' reading, mathematical and scientific competence as well as of their nonverbal cognitive basic skills (perceptual speed and reasoning).[10] The NEPS competence data files contain item variables (responses to the test items) as well as overall competence scores for each domain.[11] We use the latter measures in our analyses.

Given our interest in assessing the link between educational achievement and network characteristics, we focus on those students who reported that they would rather or very likely receive the respective type of support from their social network (as they were asked to provide further information on the size and structure of their network). Whereas both students from regular schools and with special educational needs are sampled in starting cohort 4, our analyses include the former only.

Descriptive statistics for all variables used in our empirical analyses are provided in Table 6 in the Appendix. This table also provides summary information for all observations except those from students attending special-needs schools.[12]

Our methodological approach for studying the impact of inconsistent answering behavior on the estimated effect of education comprises two phases:

---

[10] A detailed description of the tested constructs, test processes and sample items as well as information on the psychometric properties of the data can be found at https://www.neps-data.de/en-us/datacenter/dataanddocumentation/startingcohortgrade9/documentation.aspx. Students' reading and cognitive skills were assessed in the second data collection wave (i.e., when data on network size and structure was also collected); students' mathematics and scientific skills were assessed in the first wave. However, the time lag between the two waves is quite small. While NEPS participants are generally followed up annually, the fourth starting cohort started with two short distance waves.

[11] Weighted maximum likelihood estimates in tests that are scaled based on models of item response theory (i.e., reading, mathematics and scientific competence) and sum scores to measure perceptual speed and reasoning. Detailed information on how the scores were constructed is provided in the technical report by Pohl & Carstensen (2012).

[12] The actual numbers of observations used in specific empirical models might be smaller due to missing values in some of the explanatory and/or explained variables and are provided in the respective step of the analysis.

In the first phase, we aim at providing insight into whether inconsistent response style systematically depends on which type of school students attend and on whether they have a migration background or not (Section 3). To this end, a technical operationalization of "inconsistent answering behavior" is provided, which will be employed in the further empirical analyses. Based on this operationalization, we first present descriptive evidence on how prevalent inconsistent answering behavior is across different subgroups of students. In particular, we break down the numbers by type of school and migration background. This descriptive evidence is converted in a statistical model of answering behavior by means of a probit specification, which allows conclusions on the partial marginal effect of factors on students' answering behavior. Apart from our focus on the type of school attended, we also consider measures of competencies available in the dataset as further measures of educational achievement.

In the second phase (Section 4) we estimate the association between educational achievement and network characteristics. Given our research interest in the investigation of non-monetary returns to education in the form of social inclusion, in this article we estimate models to explain the variation in level of education and migration background of students' networks as a function of students' educational achievement, specifically the type of school they attend. We compare the estimates derived from a naïve model in which inconsistent cases are completely ignored with the estimates derived from a model in which we employ classical Heckman correction, which – under certain assumptions – should eliminate the bias of excluding inconsistent cases by means of taking the selection process explicitly into account. Thereby, we concentrate on modeling the impact of the type of school attended on social network characteristics. Measures of students' competencies are used in the selection equation only as this type of information is not included in many datasets.

To reduce the endogeneity bias that may result from the fact that educational achievement (which is an explanatory variable in our models) is likely to be influenced by network characteristics (which is the dependent variable in our models) itself, we focus on explaining the share of contacts with a migration background who could provide help in terms of providing information about interesting VET positions (as one out of the four sets of questions asked to assess students' social resources) to illustrate the potential selection bias. In this case, the assumption that the dependent variable is in fact caused by the type of school attended is more plausible than in case of the other questions. For instance, schoolmates will be more likely to be able to provide information about interesting VET positions than to eventually get a VET position for the target person. To take into consideration that migrants have a higher probability of knowing migrants who could provide help no matter what type of school they attend, we control for the target person's migration background. This way, we also consider that migrants may be more likely to make response errors due to potential language barriers.

3. Evaluation of answering behavior

This section provides information about the extent to which respondents fail to provide consistent answers to the previously described questions to assess network size and structure.

3.1. Operationalization of inconsistent responses

The first step of the analysis consists in measuring the share of respondents who failed to provide answers in accordance with the structure of the questionnaire. We focus on the information provided about the number of persons who would potentially provide support as well as on these persons' educational and migration background (see question number 22, Figure 1).

According to the questionnaire, the column in which the respondent is supposed to report the network's share with the highest level of general education and with a migration background depends on the respondent-reported network size (i.e., one, two or three or more persons). We consider answers as consistent if the respondent (1) reported the network size and (2) completed the question on the share (2a) who has completed the highest level of general education and (2b) with a migration background in the correct column, whereby we report separate error rates for each of the two network characteristics. We consider answers as inconsistent if the respondent checked answers in the wrong column or in more than one column. Observations with no information on network size and/or on the respective network characteristic are set to missing. The indicator variables take on the value 1 if inconsistent answers are observed and 0 for consistent answers for both types of support (i.e. *info jobs* and *help jobs*).

For the sample split according to students' migration background, we define a variable *migb* which takes on the value 0 if the student and both parents are born in Germany and 1 if the student or at least one parent was born abroad.

Table 6 in the Appendix provides a list of all variables used in our empirical analysis including descriptive statistics for our samples, i.e. for students who provided answers to the *info jobs* and *help jobs* questions, respectively.

3.2 Descriptive evidence

Descriptive evidence on the occurrence of inconsistent answering behavior regarding network characteristics of *info jobs* and *help jobs* is provided in Table 1 and Table 2, respectively. The two blocks of columns correspond the share of persons with a migration background and to the share of persons with *Abitur*. Within each block, the results are reported separately for students without and with a migration background (*migb*). Rows correspond to the attended type of school. For each population subgroup, the first column provides the share of inconsistent answers (in percent); the second one shows the size of the respective group of students.

info jobs	network members with migration background (*share_migb*)				network members with Abitur (*share_abi*)			
	migb = 0	N	migb = 1	N	migb = 0	N	migb = 1	N
Hauptschule	14.71	1,285	20.29	828	13.86	1,255	19.66	814
Realschule	5.76	1,667	10.87	469	4.47	1,634	7.83	460
Gymnasium	2.44	2,993	3.67	627	1.26	2,948	3.40	617
Comprehensive School	8.59	594	10.31	97	6.77	576	9.28	97
Integrated School	7.81	768	9.78	317	6.58	760	9.55	314
Total	6.42	7,307	12.10	2,338	5.20	7,173	11.12	2,302

Table 1: Share of inconsistent responses regarding network structure for information about VET positions (NEPS, own calculations).

help jobs	network members with migration background (*share_migb*)				network members with Abitur (*share_abi*)			
	migb = 0	N	migb = 1	N	migb = 0	N	migb = 1	N
Hauptschule	9.19	1,436	11.23	873	8.80	1,386	11.18	859
Realschule	3.12	1,826	4.82	519	2.35	1,789	3.19	502
Gymnasium	1.36	3,317	1.16	688	0.89	3,275	1.18	679
Comprehensive School	4.44	675	7.07	99	3.04	658	5.10	98
Integrated School	2.97	841	7.19	334	2.18	826	8.28	338
Total	3.57	8,095	6.45	2,513	2.91	7,934	6.18	2,476

Table 2: Share of inconsistent responses regarding network structure for effort to find a VET position (NEPS, own calculations).

A first observation is that a substantial share of students did not provide consistent answers. Also, as expected, the data clearly points to a much higher error rate in the population of students with a migration background, and to a systematic variation in the error rates across the different types of school: Across the two sets of questions (i.e., *info jobs* and *help jobs*) and the two dimensions of network structure (i.e., share with a migration background and share with *Abitur*), the error rates range from less than one up to about 15 percent for students without a migration background, and from about 1.2 to as much as 20 percent for students with a migration background.

The fraction of inconsistent answers is much smaller in the second set of questions (*help jobs*). Also, the number of students who completed the second set of questions exceeds the number who completed the first set of questions. This observation suggests that the lower error rate in the second set of questions may not be explained by the fact that students with lower abilities stop answering the questions on their social resources after having completed the first set of questions comparatively often but rather points to learning effects.

The highest shares of inconsistent answers are observed for students attending the lowest track (*Hauptschule*) with values up to about 15 percent for students without a migration background and up to as much as 20 percent for students with a migration background. The shares decrease when moving on to students attending the intermediate track (*Realschule*). As expected, the smallest values are observed for students attending the highest educational track (*Gymnasium*). The error rates for students attending schools that offer several tracks range between the shares observed for students attending *Hauptschule* and *Realschule*. Interestingly, the differences between students with and without a migration background decrease from less to more demanding educational tracks but persist. An exception is the group attending the most demanding type of school (*Gymnasium*) for *help jobs*.

3.3 Modeling answering behavior

In order to establish a better understanding of inconsistent answering behavior, we estimate probit models for binary dependent variables. As dependent variables we use the indicator variables introduced in Section 3.1, which take on the value 1 for inconsistent answers and 0 else. The indicator variable for the question about information on VET positions is labeled *ind_info*; the variable for the question about actual support with finding a job is labeled *ind_help*. We differentiate whether the inconsistency arises in the part of the question to assess the network share with a migration background (*share_migb*) or in the part addressing the network's educational background (*share_abi*).

We consider the following factors as predictors for the likelihood of providing inconsistent answers: Given the complexity of the questions, we assume that individual competencies might have an impact on the probability of providing inconsistent answers. Therefore, we include information on the type of school attended (reference: *Hauptschule*) as well as measures of students' reading skills (*read*), mathematical competence (*math*) and scientific competence (*science*) and measures of specific cognitive skills (perceptional speed (*speed*) and reasoning (*reas*)). For each domain, we use the summary scores provided by the NEPS (Skope, Pink and Bela, 2013; Pohl and Carstensen, 2012) described in section 3.1. In consideration that migrants may be more likely to make response errors than natives due to migration-specific conditions such as language barriers, we further include information on students' migration background (*migb*). In addition, we control for sex (*sex* = 2 for females).

The results of the probit estimations are summarized in Table 3 for *ind_info* and in Table 4 for *ind_help*.

ind_info	network members with migration background (share_*migb*)				network members with Abitur (share_*abi*)			
	Coef.	Std.Err.	z	p	Coef.	Std.Err.	z	p
Competencies								
read	-0.1569	0.0248	-6.32	0.000	-0.1743	0.0273	-6.39	0.000
math	-0.0751	0.0283	-2.65	0.008	-0.0431	0.0316	-1.37	0.172
science	-0.1449	0.0345	-4.20	0.000	-0.1943	0.0384	-5.06	0.000
speed	-0.0001	0.0014	-0.09	0.925	0.0002	0.0015	0.15	0.880
reas	-0.0468	0.0094	-4.96	0.000	-0.0621	0.0101	-6.16	0.000
Type of school								
Realschule	-0.1797	0.0600	-2.99	0.003	-0.2488	0.0653	-3.81	0.000
Gymnasium	-0.2567	0.7257	-3.54	0.000	-0.3638	0.0820	-4.44	0.000
Comprehensive School	-0.2146	0.0827	-2.59	0.010	-0.2686	0.0894	-3.00	0.000
Integrated School	-0.1953	0.0704	-2.77	0.006	-0.2047	0.0741	-2.76	0.006
Other factors								
migb	0.0785	0.0470	1.67	0.095	0.0972	0.0506	1.92	0.055
sex	-0.2992	0.0452	-6.62	0.000	-0.3167	0.0496	-6.38	0.000
Constant	-0.8504	0.1142	-7.45	0.000	-0.8442	0.1227	-6.88	0.000
McFadden Pseudo R^2				0.1313				0.1742

Table 3: Probit model for inconsistent responding regarding network structure for information regarding apprentice positions (NEPS, own calculations).

ind_help	network members with migration background (share_migb)				network members with Abitur (share_abi)			
	Coef.	Std.Err.	Z	p	Coef.	Std.Err.	Z	p
Competencies								
read	-0.1402	0.0293	-4.79	0.000	-0.2062	0.0323	-6.39	0.000
math	-0.0468	0.0335	-1.39	0.163	0.0073	0.0366	0.20	0.842
science	-0.1866	0.0405	-4.61	0.000	-0.1783	0.0448	-3.98	0.000
speed	-0.0018	0.0017	-1.08	0.280	-0.0013	0.0018	-0.74	0.460
reas	-0.5920	0.0108	-5.49	0.000	-0.0805	0.0115	-6.96	0.000
Type of school								
Realschule	-0.1648	0.0700	-2.35	0.019	-0.2644	0.0776	-3.41	0.001
Gymnasium	-0.2211	0.0860	-2.57	0.010	-0.2701	0.0959	-2.82	0.005
Comprehensive School	-0.2042	0.0947	-2.16	0.031	-0.3446	0.1091	-3.16	0.002
Integrated School	-0.2135	0.0828	-2.58	0.010	-0.1718	0.0859	-2.11	0.034
Other factors								
sex	-0.3778	0.0535	-7.06	0.000	-0.3384	0.0585	-5.78	0.000
migb	0.0271	0.0553	0.49	0.624	0.0687	0.0590	1.16	0.245
Constant	-0.9487	0.1314	-7.22	0.000	-0.9221	0.1396	-6.61	0.000
McFadden Pseudo R^2				0.1429				0.1841

Table 4: Probit model for inconsistent responding regarding network structure for support regarding apprentice positions (NEPS, own calculations).

The results clearly demonstrate that the probability of providing consistent answers to the complex questions about network size and structure depend on individual-specific characteristics to a substantial extent. Focusing first on the measures of specific competencies, it becomes obvious that higher reading competencies (*read*) go hand in hand with a reduced risk of inconsistent answers. The same applies to scientific competencies (*science*) and the cognitive competency of reasoning (*reas*). These variables are significant (at least at the 1% level) in all four settings. In contrast, the coefficient for mathematical competencies (*math*) is significant at the 1% level in one out of the four settings (*ind_info*, share with migration background) but not significant in any of the other three settings. Neither does the speed of perception (*speed*) exhibit a significant influence.

Remarkably, the type of school attended has an independent effect on the probability of the occurrence of inconsistent answers (i.e., when individual competences are controlled for). Attending any type of school different from *Hauptschule* decreases the risk of providing inconsistent answers in a statistically significant way (at least at the 5% level). Having a migration background (*migb*), in contrast, appears to increase the risk of inconsistent answers only slightly when competence measures and information on the type of school attended are controlled for. The effect is not significant in *ind_help*, and significant at the 10% level only in *ind_info*. Finally, we find that female respondents (*sex*) have a lower risk of providing inconsistent answers. This effect is statistically significant at the 1% level for all four cases.

Table 7 in the Appendix provides the average marginal effects corresponding to the estimates displayed in tables 3 and 4. They show that the probability of providing inconsistent answers shrinks by 2.5 to 4 percentage points for students who attend a school type other than *Hauptschule* for *ind_info*. The marginal effects are slightly smaller for *ind_help* (1.5 to 2.5 percentage points).

Regarding the competence measures, the strongest effects are found for *read*, *science*, and *reas*. A one standard deviation increase in these competencies reduces the probability of the occurrence of inconsistent answers by about 1 to 2 percentage points. The marginal effects for *math* and *speed* are of negligible size. The probability of providing inconsistent answers is about three percentage points higher for males than for females.

4. Potential effects of misreporting

In this section we empirically illustrate the potential bias caused by neglecting inconsistent answering behavior when estimating the association between educational achievement on network characteristics. As an example, we estimate the effect of educational achievement on the share of migrants out of those persons who would provide information about interesting VET positions. As discussed above, we are specifically interested in the stability of the effect of the type of school variable.

We first estimate a simple OLS-model to explain the share of network members with a migration background. In this model, only observations with consistent answers are considered. The outcome variable is treated as a metric variable, although due to the specific data collection procedure it can take on only a limited number of values. The explanatory variables in the models are the respondent's sex (*sex*), migration background (*migb*) and dummy variables for the type of school attended (*Hauptschule* being the reference category as in the analyses above). As mentioned above, we do not include competence measures as these are not available in many datasets.

As a comparison, we estimate a Heckman model to explicitly model inconsistent answering behavior. In the first step, a probit model is used to determine the probability of providing inconsistent answers (results are shown in Table 3). The bias that may result from neglecting this non-random selection is avoided (or at least reduced) in the second stage of the Heckman procedure by adding a regressor to the model which is a function of the estimated probability of providing consistent answers (Mill's ratio).

The results are displayed in Table 5. The first column (Naïve model) shows the OLS-estimates when non-random selection is ignored. The second column provides the results of the second step of the Heckman two-stage estimator.

Info_jobs, share_migb	Naïve model (N = 8881)				Heckit (two step) (N = 9122/8416)			
	Coef.	Std.Err.	t	p	Coef.	Std.Err.	z	p
Type of school								
Realschule	-0.2659	0.045	-5.8	0.000	0.0710	0.095	0.74	0.457
Gymnasium	-0.4012	0.041	-9.71	0.000	0.1026	0.102	1.00	0.318
Comprehensive School	-0.3923	0.065	-6.03	0.000	-0.1328	0.123	-1.07	0.284
Integrated School	-0.1904	0.055	-3.45	0.001	0.1217	0.108	1.12	0.318
Other factors								
sex	0.0532	0.029	1.8	0.072	0.1541	0.056	2.75	0.006
migb	2.7682	0.356	77.7	0.000	2.6347	0.066	39.3	0.000
Constant	1.6852			0.000	0.9878	0.119	8.28	0.000
R^2				0.4299				n.a.

Table 5: Results from different models for share_migb.

With the exception of the coefficient estimated for migration background, the results reveal a substantial variation of the order of magnitude and the statistical significance of coefficients for the explanatory variables across the two models. The effect of students' sex, which is not statistically significant in the naïve model, triplicates in size and becomes statistically significant at the 1% level when taking into account the selection process leading to inconsistent answers. The significant effects of the school variables in the naïve model virtually disappear. In the naïve model, attending any type of school other than *Hauptschule* significantly decreases the share of migrants who could provide information about interesting VET-positions. All coefficients are significant at least at the 1% level. None of the coefficients is significant at any conventional level in the model that takes into account selection processes. Also, three out of four coefficients change sign. Solely, the positive and rather large coefficient for the migration background dummy is quite stable and significant at the highest level when using a misspecified model, indicating that the share of migrants who could provide information about interesting VET-positions is significantly higher in migrants' networks than in natives' networks.

Note that the observations included in the naïve model and the second step of the Heckit specification differ due to missing observations for the additional variables on individual competences and cognitive skills included in the selection equation of the Heckit model. We repeated the analysis for the naïve model including exactly the same observations as used for the Heckit model (Table 8 in the appendix, column 1). The coefficients are stable both with regard to size and significance. Furthermore, to rule out the possibility that the significance of the school type dummies in the naïve model results from the missing covariates on individual competences and cognitive skills used in the first step of the Heckit model, we also repeated the analysis including these variables as additional regressors (see Table 8 in the appendix, column 2). Scientific and reading competence are found to have a significant effect (at the highest level) even when school type is controlled for. No significant effects are found for mathematical competence and cognitive skills. The coefficients estimated for the school type dummies decrease in size but remain significant, even though at a lower level for *Gymnasium* and *Realschule*. The only exception is *Integrated School*, which is not significant anymore when individual competences and cognitive abilities are controlled for.

5. Discussion and concluding remarks

Summarizing, we find that a substantial share of students was unable to answer questions regarding their social networks consistently. Also, we found the error rate to vary systematically across the different types of school attended, and to be substantially higher in the population of students with a migration background (also within the same type of school). Only among students attending the most demanding type of school (*Gymnasium*) migration background has only a limited impact on the error rate.

Furthermore, the two questions regarding *info jobs* and *help jobs* immediately follow one another. As might have been expected, the descriptive evidence suggests an immediate learning effect in that the share of inconsistent answers for *help jobs* (Table 2) is smaller than for *info jobs* (Table 1). An alternative explanation for the lower incidence of inconsistent answering behavior might be the overall smaller network size when it comes to making actual effort to get a VET position for the target person (as compared to the mere provision of information). For instance, when the student only knows one or two persons, it might be easier to indicate whether this person has a migration background or not as compared to the situation when a summary statement has to be made about several persons. To check whether our findings depend on the specific choice of indicators for inconsistent behavior, we also studied alternative operationalizations. For example, we considered multiple answers across columns as consistent if they correspond to the same share. None of these alternative specifications lead to results that significantly differ from the results presented above.

The descriptive results are supported by the results from the probit model, which show that the probability of providing consistent answers to the complex questions about network characteristics significantly depends on individual-specific characteristics. Attending any type of school different from *Hauptschule* reduces the risk of providing inconsistent answers, even when individual competencies are controlled for. Females have a significantly lower probability to inconsistently respond, whereas no systematic differences are observed between students with and without a migration background when type of school and individual competencies are controlled for.

The comparison of the estimates derived from a model in which inconsistent answers are simply ignored with the estimates derived from a Heckit model suggests that ignoring the selection due to inconsistent answers affects inference on the effect of education on network characteristics to a relevant extent: The results from the naïve model just leaving out inconsistent answers suggest that the share of network members with a migration background depends in a statistically significant way on the type of school attended, i.e. this share shrinks when moving from *Hauptschule* to *Realschule* or *Gymnasium* all else equal. However, after applying a standard Heckit model, it turns out that these effects might have to be considered as artefacts stemming from inconsistent answers, which do not come up randomly, but are a function of type of school attended, but even more so of specific competencies acquired by the respondents. Taking these effects into account, we find that the share of network members with migration background depends in a statistically significant way on sex and own migration background, but not on the type of school attended.

It has to be stressed that the NEPS just serves as an example, and that it cannot be concluded from our results that the same inconsistencies arise in data sets using different questions on different participants. However, our results clearly underline the statement made by Marsden (2014, p. 377): "In any event, the visual design of self-administered instruments warrants care, […]", suggesting that the use of less complex questions might be required to arrive at lower inconsistency rates and to avoid misleading inference with respect to the effect and determinants of students' social networks. At any rate, even though the above estimations illustrate the potential bias that might result from the neglection of non-random variation due to inconsistent response style only, they clearly point to the need to check and, if detected, correct for non-random selection when analyzing the effect or determinants of the social resources students have at their disposal in more detail.

References:

Baumert, J./Stanat, P./Watermann, R. (2006). „Schulstruktur und die Entstehung Differenzieller Lern- und Entwicklungsmilieus“. In: J. Baumert/P. Stanat/R. Watermann (Ed.): Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 95-188.

Becker, B./Gresch, C. (2016). „Bildungsaspirationen in Familien mit Migrationshintergrund“. In: C. Diehl/C. Kristen/C. Hunkler (Ed.): Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 73-115.

Becker, B. (2010). „Bildungsaspirationen von Migranten: Determinanten und Umsetzung in Bildungsergebnisse“. MZES Working Paper No. 137. Mannheim.

Blossfeld, H.-P./Roßbach, H.-G./von Maurice, J. (Ed.) (2011). “Special Issue on Education as a Lifelong Process: The German National Educational Panel Study (NEPS)”. Zeitschrift für Erziehungswissenschaft: 14. Wiesbaden: VS Verlag für Sozialwissenschaften.

Bradburn, N. (1979). Respondent Burden. In: Reeder, L. (Ed.): Health Survey Research Methods: Second Biennial Conference, Williamsburg, VA. Washington, DC: U.S. Government Printing Office.

Bygren, M./Szulkin, R. (2010). “Ethnic Environment During Childhood and the Educational Attainment of Immigrant Children in Sweden”. Social Forces 88(3) pp. 1305-1329.

Fuß, D./von Maurice, J./H.-G. Roßbach (2016). „A Unique Research Data Infrastructure for Educational Research and Beyond: The National Educational Panel Study”. Journal of Economics and Statistics 236(4) pp. 517–528.

Kroneberg, C. (2008). “Ethnic Communities and School Performance Among the New Second Generation in the United States: Testing the Theory of Segmented Assimilation”. The Annals of the American Academy of Political and Social Science 620(1) pp. 138-160.

Manski, C.F. (2004). “Measuring Expectations”. Econometrica 72(5) pp. 1329-1376.

Pohl, S./C.H. Carstensen (2012). “NEPS Report – Scaling the Data of the Competence Tests”. NEPS Working Paper No. 14. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Marsden, P.V. (2014). “Survey Methods for Network Data”. In: J. Scott/ P.J. Carrington (Ed.): The SAGE Handbook of Social Network Analysis. London: SAGE Publications pp. 370-388.

Morgan, S.L. (1998). “Adolescent Educational Expectations: Rationalized, Fantasized, or Both?”. Rationality and Society 10(2) pp. 131-162.

Morgan, S.L. (2002). “Modeling Preparatory Commitment and Non-Repeatable Decisions: Information-Processing, Preference Formation and Educational Attainment”. Rationality and Society, 14(4) pp. 387-429.

Roth, T./Salikutluk, Z./Kogan, I. (2010). “Auf die „richtigen “Kontakte kommt es an! Soziale Ressourcen und die Bildungsaspirationen der Mütter von Haupt-, Real-und Gesamtschülern in Deutschland“. In: Becker,B./Reimer, D. (Ed.): Vom Kindergarten bis zur Hochschule - Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 179-212.

Skopek, J./Pink, S./Bela, D. (2013). „Starting Cohort 4: 9th grade (SC4). SUF Version 1.1.0. Data Manual”. NEPS Research Data Paper. Bamberg: National Educational Panel Study (NEPS), University of Bamberg.

Stubbe, T.C. (2009). „Bildungsentscheidungen und Sekundäre Herkunftseffekte: Soziale Disparitäten bei Hamburger Schülerinnen und Schülern der Sekundarstufe I“. Empirische Erziehungswissenschaft 14. Münster: Waxmann.

Stubbe, T.C./Bos W./ Euen, B. (2012). „Der Übergang von der Primar-in die Sekundarstufe“. In Bos, W./Tarelli, I./Bremerich-Vos, A./Schwippert, K. (Ed.): IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im Internationalen Vergleich. Münster: Waxmann. pp. 210-226.

Trebbels, M. (2014). “The Transition at the End of Compulsory Full-time Education: Educational and Future Career Aspirations of Native and Migrant Students”. Wiesbaden: VS Verlag für Sozialwissenschaften.

Von Maurice, J./Leopold, T./Blossfeld, H.-P. (2011). “The National Educational Panel Study: A Long-term Assessment of Competence Development and Educational Careers”. NEPS Working Paper No. 2. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Von Maurice, J./Sixt, M./Blossfeld, H.-P. (2011). “The German National Educational Panel Study: Surveying a Cohort of 9th Graders in Germany”. NEPS Working Paper No. 3. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Wagner, G.G./Frick, J.R./Schupp, J. (2007). “The German Socio-Economic Panel Study (SOEP) – Evolution, Scope and Enhancements”. Schmollers Jahrbuch 127 pp. 139-169.

Winker, P. (2016). “Assuring the Quality of Survey Data: Incentives, Detection and Documentation of Deviant Behavior”. Statistical Journal of the IAOS 32 pp. 295-303.

Appendix

Descriptive Statistics

Variable	Obs.	Mean	Std.dev.	Min	Max
Inconsistent answers					
Share Abitur: *info jobs*	8957	0.065	0.247	0	1
Share Abitur: *help jobs*	9839	0.037	0.189	0	1
Share migrants: *info jobs*	9122	0.077	0.267	0	1
Share migrants: *help jobs*	10034	0.043	0.202	0	1
Share with Abitur/migration background					
Share Abitur: *info jobs*	8372	3.645	2.102	1	7
Share Abitur: *help jobs*	9476	2.989	1.677	1	7
Share migrants: *info jobs*	8416	2.079	1.835	1	7
Share migrants: *help jobs*	9605	2.133	1.286	1	7
Competencies					
Scientific competence (*science*)	11157	0.069	0.991	-3.558	5.287
Mathematical competence (*math*)	11157	0.091	1.216	-4.370	4.619
Reading competence (*read*)	11157	0.072	1.236	-4.746	3.300
Perceptual speed (*speed*)	11157	59.305	13.726	2	93
Reasoning (*reas*)	11157	8.770	2.396	0	12
Type of school					
Hauptschule	11157	0.219	0.414	0	1
Realschule	11157	0.222	0.416	0	1
Gymnasium	11157	0.375	0.484	0	1
Comprehensive School	11157	0.072	0.259	0	1
Integrated School	11157	0.111	0.315	0	1
Other factors					
Migration background (=1) (*migb*)	11157	0.243	0.429	0	1
Female (=2) (*sex*)	11157	1.516	0.410	1	2

Table 6: Descriptive Statistics

Average Marginal Effects

	network members with migration background (share_*migb*)				network members with Abitur (share_*abi*)			
	ind_info		*ind_help*		*ind_info*		*ind_help*	
	AME	p	AME	p	AME	p	AME	p
Competencies								
read	-0.0200	0.000	-0.0111	0.000	-0.0187	0.000	-0.0140	0.000
math	-0.0096	0.008	-0.0037	0.163	-0.0046	0.172	0.0000	0.842
science	-0.0185	0.000	-0.0148	0.000	-0.0208	0.000	-0.0121	0.000
speed	-0.0000	0.925	-0.0001	0.280	0.0000	0.880	-0.0000	0.460
reas	-0.0060	0.000	-0.0047	0.000	-0.0067	0.000	-0.0054	0.000
Type of school								
Realschule	-0.0249	0.003	-0.0142	0.018	-0.0294	0.000	-0.0190	0.000
Gymnasium	-0.0339	0.000	-0.0182	0.008	-0.0399	0.000	-0.0193	0.003
Comprehensive School	-0.0291	0.005	-0.0171	0.019	-0.0313	0.001	-0.0233	0.000
Integrated School	-0.0268	0.004	-0.0177	0.006	-0.0248	0.004	-0.0138	0.026
Other factors								
migb	0.0102	0.102	0.0022	0.627	0.0107	0.061	0.0047	0.254
sex	-0.0382	0.000	-0.0299	0.000	-0.0340	0.000	-0.0227	0.000

Table 7: Average Marginal Effects for Probit model estimates shown in Table 3 and Table 4.

Info_jobs, share_migb	Naïve model, same obs. as Heckit (N = 8416)				Naïve model, additional controls (N = 8416)			
	Coef.	Std.Err.	t	p	Coef.	Std.Err.	t	p
Competencies								
read	-	-	-	-	-0.071	0.017	-4.18	0.000
math	-	-	-	-	-0.025	0.018	-1.39	0.164
science	-	-	-	-	-0.096	0.023	-4.2	0.000
speed	-	-	-	-	0.002	0.001	1.8	0.072
reas	-	-	-	-	-0.014	0.007	-1.9	0.057
Type of school								
Realschule	-0.283	0.047	-6.00	0.000	-0.147	0.049	-3.01	0.003
Gymnasium	-0.418	0.042	-9.82	0.000	-0.123	0.052	-2.35	0.019
Comprehensive School	-0.404	0.067	-5.98	0.000	-0.343	0.067	-5.07	0.000
Integrated School	-0.2	0.056	-3.52	0.000	-0.088	0.057	-1.54	0.125
Other factors								
sex	0.034	0.03	1.15	0.252	-0.008	0.032	-0.25	0.803
migb	2.772	0.036	75.7	0.000	2.692	0.037	72.3	0.000
Constant	1.709	0.039	43.1	0.000	1.622	0.094	17.2	0.000
R^2				0.43				0.437

Table 8: Results from different models for share_migb with reduced sample and additional controls.