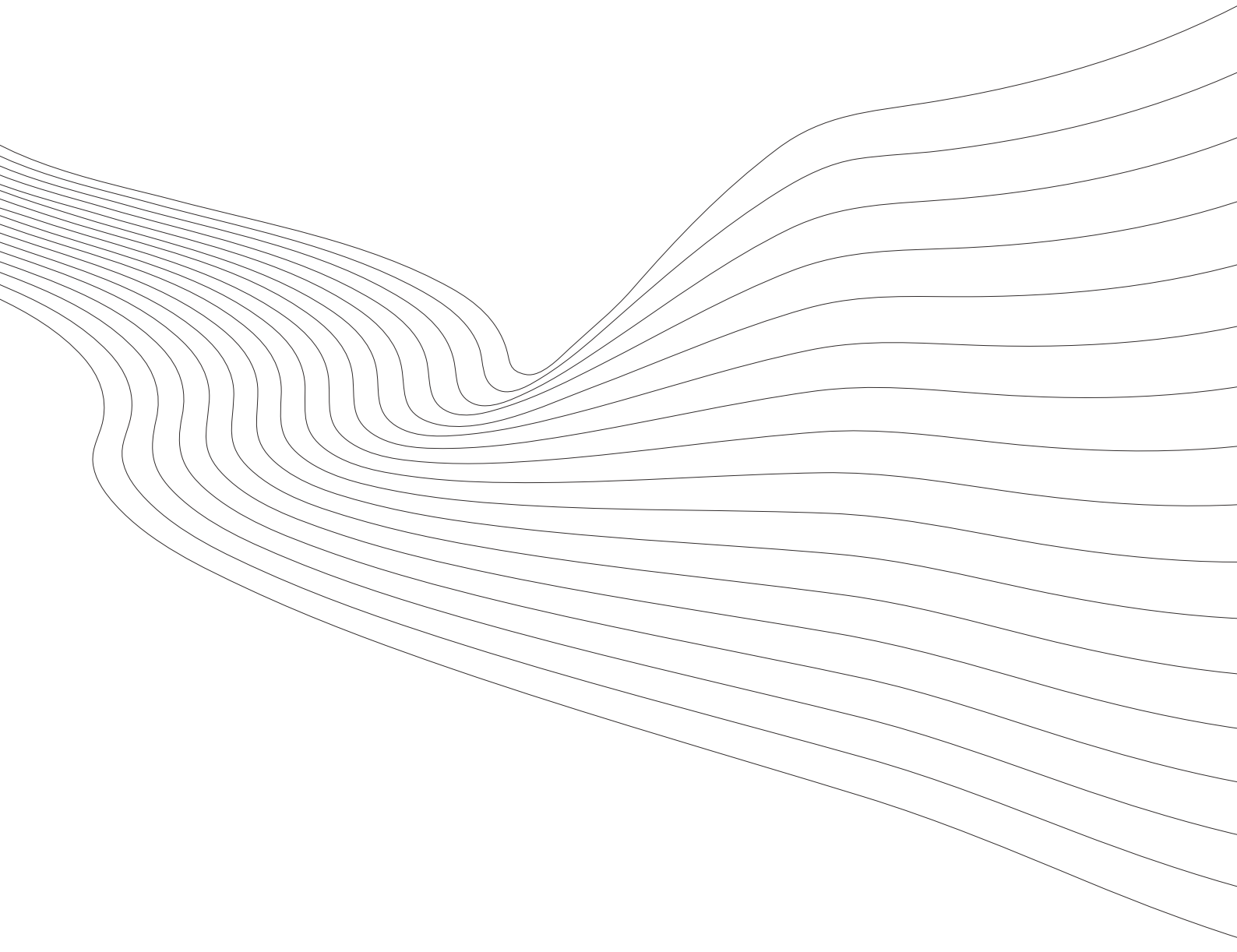
KOF

KOF Working Papers

Can a Standardized Aptitude Test Predict Training Success of Apprentices? Evidence from a Case Study in Switzerland

Michael Siegenthaler



No. 270
February 2011

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

KOF

ETH Zurich
KOF Swiss Economic Institute
WEH D 4
Weinbergstrasse 35
8092 Zurich
Switzerland

Phone +41 44 632 42 39
Fax +41 44 632 12 18
www.kof.ethz.ch
kof@kof.ethz.ch

Can a standardized aptitude test predict training success of apprentices? Evidence from a case study in Switzerland

Short title: Aptitude tests and training success

Michael Siegenthaler[1]

February 2011

Abstract

Due to a widely spread distrust in the signalling value of school grades, Swiss employers require external, standardized aptitude test results when recruiting new apprentices. However, the predictive quality of such test results has never been thoroughly researched. Therefore, this case study analyses whether external aptitude tests can improve the quality of predicting success in apprenticeship training. I find that such information is a) not correlated with school grades at the end of compulsory schooling but b) does not add information that would explain either the success in VET schooling (school grades in the first and second year of apprenticeship training), the probability of unexcused vocational school absences or the likelihood of a premature ending of the apprenticeship contract.

Keywords: Apprenticeship, hiring, aptitude test, predictive validity, screening

JEL classifications: I21, J24, M51, M53

[1] Address for correspondence: Swiss Economic Institute (KOF), Weinbergstrasse 35, CH-8092 Zurich. E-mail: siegenthaler@kof.ethz.ch. Phone: +41 44 633 93 67. I thank Stefan Wolter for his assistance during the whole research process and for his comments on my Master thesis. I am grateful for the support by the firm Multicheck and I thank Ralph Hardegger, Michèle Oswald, Bernd Schauer, Simon Siegenthaler, and participants at the Brown Bag seminar at ETH Zurich for helpful comments and suggestions.

1. Introduction

During the last fifteen years, standardized aptitude tests for apprenticeships, provided by private firms, have increasingly influenced adolescents' transition from lower secondary school into dual training programmes in Switzerland.[2] Nowadays, many enterprises that hire apprentices require such a test from their applicants, particularly in the commercial sector or the retail business. This study examines the predictive power of one such test: the "Multicheck retail sale". Is the test a valid predictor of vocational success of apprentices? By answering this question, we also gain insights whether the test can improve the apprentice selection of a firm.

The success of aptitude tests for apprenticeships is closely linked to the deficiencies of performance measures provided by lower secondary school (Imdorf 2009; Moser 2004). Employers require the tests because they cannot infer the true scholastic abilities and cognitive skills of their candidates from school grades and the type of school the applicant attended. Employers' mistrust in school grades and may be appropriate. School achievements are a noisy and biased measure of effective scholastic abilities of pupils, rendering them nearly uncorrelated to "objective" measures of scholastic and cognitive ability such as PISA scores or IQ tests (Baron-Boldt 1989; Bauer and Sheldon 2008; Bishop 1994; De Paola and Scoppa 2010; Kronig 2007; Lindahl 2007). Moreover, the Swiss school system is very opaque. Cantons and sometimes even municipalities differ considerably in the amount of educational tracking and curricula, for example. School reforms have increased the heterogeneity among school leavers. Comparing academic achievements of applicants for an apprenticeship position entails substantial problems for the firm.

[2] Upper secondary education in Switzerland is to a great extent characterized by dual training programmes: firm-based provision of training combined with theoretical education in vocational schools. See, e.g., Mühlemann et al. (2009) for a description of the dual apprenticeship system in Switzerland.

The tests seem to alleviate this informational deficit. Their standardization enables comparison of scholastic and cognitive abilities across candidates in a simple and economic way. The tests economize candidate selection. It is hence not surprising that studies have found a considerable impact of the test results on the chances of being employed as an apprentice (Moser 2004; Mühlemann et al. 2007).

In improving the "signalling" of scholastic and cognitive competences of pupils to employers (Bishop 1994), the tests could improve the matching between applicant and firm, and hence increase social welfare (Costrell 1994). But can the tests keep their promises? What is their effective value as predictor of vocational success? Very few authors have tried to answer these questions. To my knowledge, the only scientific attempt for a Swiss aptitude test is a study by Widmer (2006) examining a test for the commercial sector.[3]

This paper contributes to fill this gap. It assesses the predictive power of the "Multicheck retail sale" *(Multicheck Detailhandel und Service)*. All major Swiss enterprises in the retail business require the results from this test from their applicants to an apprenticeship. The informational value of the test is examined using a self-collected data set containing information about personal status, lower secondary school reports, Multicheck results and vocational school reports of 334 apprentices, employed at *Migros*, the biggest employer in the retail business in Switzerland. The main outcome variable considered is vocational school grades in the first and third semester. The results are verified using the probability of unexcused vocational school absences and of premature apprenticeship terminations as additional criteria of vocational success.

The regressions demonstrate a poor performance of the overall score (*Gesamtergebnis*) of the Multicheck in predicting vocational school performance in both semesters considered. School grades and the type of school the applicant attended are

[3] However, the study suffers partly from small sample sizes. Investigations of Multicheck (2010) about the validity of their tests do not meet scientific requirements since they lack documentation.

noticeably more reliable sources of information for the recruiter. These findings are robust to different specifications and estimation methods. Additionally, they carry over to the other outcome variables considered.

The structure of the paper is as follows. In the next section, I discuss the informational value that can be theoretically expected from standardized aptitude tests for apprenticeships. Section 3 provides information on the data and the sample. The following sections present the main regression results (section 4) and their robustness (section 5). The paper is concluded after the discussion of the two additional outcome variables (section 6).

2. The Multicheck and the prediction of training success

2.1 The Multicheck retail sale

Aptitude tests for apprenticeships are provided and carried out by specialized private firms in Switzerland. "Multicheck" and "basic-check" are the most widely used tests. Even small and medium-sized enterprises require them from their applicants. By its own account, the firm Multicheck is the biggest provider of standardized aptitude tests in Switzerland. More than 30'000 Swiss adolescents pass one of the six different "Multicheck Junior" tests each year. Multicheck calls its tests "job relevant aptitude assessments."[4] According to Widmer (2006, p. 49), their aim is "to predict practical and especially theoretical success in vocational training of apprentices."

The specific version of the Multicheck studied in this paper is the "Multicheck retail sale" (*Multicheck Detailhandel and Service*). The test is at the present time mandatory for nearly all adolescents applying for an apprenticeship in the retail business. Evaluations of the test can be obtained for different apprenticeships in the field. This study concentrates on the version for potential "retail sales specialists" (*Detailhandelsfachmann/-fachfrau*). About

[4] The expression stems from the homepage of the enterprise. See www.multicheck.ch/en/mc-junior.

10’300 adolescents passed this version of the Multicheck in 2009. The quantitative importance of the test is the main reason why it was considered in this investigation.

The content of the test is twofold. It tests knowledge acquired at school (German, French or English or both, depending on the choice of the candidate, and calculating) and cognitive skills (logic, concentration and retentiveness). The results are presented for each of the six (or seven) test sections individually. In addition, an overall score (*Gesamtergebnis*) is built. The scale used to present the scores ranges from zero to 100. This scale is relative: 50 points in the overall score imply that 50 percent of past test takers have been better and 50 percent have been worse. The relative scale does not seem to allow for statements about a test taker’s aptitude to start a certain apprenticeship. However, the evaluation of Multicheck contains such an indication: overall test scores of less than 40 points are “insufficient”, those between 40 and 60 points are “sufficient”, and test takers with more than 60 points “exceed” the requirements of an apprenticeship (cf. Appendix A).[5]

2.2 Assessing the predictive power of the test

The firms use the test to identify the quality of their candidates. One may thus label it a “screening device” (Arrow 1973; Stiglitz 1975).[6] If we know whether the test is a valid predictor of training success of apprentices, we may gain information about the test’s “usefulness” as a screening device. The latter is given by the amount the test improves the firm’s ability to assess the chances of vocational success of an applicant. In this respect, the “informational value” of the test depends on the informational surplus it adds to the selection

[5] Although Multicheck states that the decision of accepting a candidate is left to the company, it nevertheless refers to “experience” that “has shown that candidates with an overall average below 40 percent do not complete their apprenticeship, or only with struggle” in order to justify the thresholds. The argument is not very convincing: how can Multicheck make sure that exactly 39.5 percent of the test takers do not, or “insufficiently”, fulfill the requirements of the respective apprenticeship? (Note that all of the tests of Multicheck have the same thresholds.)

[6] While firms screen, applicants signal. But the situation analysed in this paper is less closely related to the concept of signalling (see Spence 1973; Weiss 1995) than to screening. Signalling refers to a self-selective mechanism which induces job searchers to gather costly “signals” of human capital–a mechanism underlying school grades. The Multicheck is more related to screening, i.e. the firm’s assessment of applicants.

decision of the firm. If all the information provided by the test is already inferable from other sources, for instance, from school grades in the application dossier, then the test would not be valuable for the employer. The informational value of the test should thus be analysed conditional on other information available in the selection. This motivates the use of multivariate regression models.[7]

The validity of the multivariate regression approach hinges on the choice of the control variables. In particular, the choice is constrained by the recruitment practices at Migros. The three cooperatives use the Multicheck nearly exclusively for the pre-selection of applicants (similar as the firms in Moser 2004). The test plays no major role in later stages of the recruitment. Hence, the inclusion of explanatory variables that actually explain success in vocational training, but are unavailable to the firm at the time of the *pre-selection* decision, may bias the results about the informational value of the Multicheck for the firm.

Consequently, the potential right-hand variables in the regressions should stem from the application dossier, the most important source of information therein being lower secondary school records. They contain information about school grades and the type of school the applicant attended. Survey evidence and empirical studies document that school grades and the type of school have a major impact on the chances that an applicant is employed (Häberlin et al. 2004; Imdorf 2009; Moser 2004; Stalder et al. 2008). A second source of information in the application dossier is unalterable personal attributes of the candidate. Attributes that are likely to influence the apprentice selection of firms are gender, social background, cultural background such as nationality and native language as well as regional origin (Amos et al. 2009; Bertschy et al. 2008; Häberlin et al. 2004; Imdorf 2005; Moser 2004; Stalder et al. 2008).

[7] The approach chosen corresponds to the psychological concept of "incremental validity": does the Multicheck add to the prediction of a certain criterion above what can be predicted by other sources of data (Hunsley and Meyer 2003)? The approach differs from the one chosen in the studies of Widmer (2006) and Multicheck (2010), in which the "predictive validity" of the test is analysed, i.e. simple correlations between the test score and a criterion of vocational success.

What variable should we use on the left-hand side of the regression, i.e. which operationalization of vocational success is relevant in the sense that the Multicheck attempts and pretends to predict it? There are at least three reasons why the most relevant outcome variable in our context is vocational school GPA. Firstly, the test mainly aims to forecast theoretical success in vocational training (see section 2.1). Secondly, the Migros uses it nearly exclusively to assess an applicant's chances of meeting the *academic* requirements of an apprenticeship. Finally, problems in vocational school are the most important reason of premature apprenticeship terminations (Stalder and Schmid 2006).

Nevertheless, I control the robustness of the results using two additional outcome variables: the probability of unexcused vocational school absences and of a premature apprenticeship termination. Both are costly for the firm[8], interrelated to vocational school grades and the respective other outcome variable, and have a relatively close relationship to practical performances of apprentices (Imdorf 2009; Stamm et al. 2010).

The reasoning of the last paragraphs boils down to the regression model presented in section 4. What hypothesis can we make about the informational value of the Multicheck in these regressions? Firstly, the test should benefit from its standardization as this prevents errors due to relative measurement, which is an important source of noise in school grades. Secondly, it measures academic and scholastic abilities of adolescents. Studies have shown that such tests can be valid predictors of labour market and educational success.[9] Moreover, if we consider that school grades and school level are likely to only imperfectly represent actual cognitive or scholastic ability of pupils, the Multicheck should have a considerable independent informational value.

---

[8] The former because the vocational instructors of the Migros will have to discuss the reasons for the absence with the apprentice, the latter because premature apprenticeship terminations, as every turnover, are costly in terms of time and financial outlay (Mühlemann et al. 2007).

[9] For tests of academic achievements such as PISA, see, e.g., Bertschy et al. (2008) or Stalder et al. (2008), and for tests of cognitive skills such as IQ tests, cf. Grant (2007), Heckman and Vytlacil (2001), or Murnane et al. (2001).

However, two reasons could limit the informational value of the test. Firstly, Stamm et al. (2010) found that more intellectually gifted apprentices do not necessarily perform better. If these results carry over to the context analysed here, the cognitive part of the Multicheck might be a poor predictor of success in vocational training. Secondly, shortcomings in the construction of the test could undermine its informational value. In this spirit, Moser (2004) has put forward concerns about the test-theoretical quality of the aptitude tests employed in Switzerland, especially concerning their test fairness and internal coherence.

3. Data

The data set covers all apprentices from three (of ten) cooperatives of Migros, the biggest employer in the Swiss retail business with a market share of 36.8 percent in 2009. The cooperatives considered are Migros Aare, Lucerne and Zurich. All apprentices started their apprenticeship as "retail sales specialists" in August 2007 or August 2008. I gathered the data directly from the personnel files of the respective cooperatives. The sample consists of 334 apprentices: 142 from Migros Aare, 84 from Migros Lucerne and 108 from Migros Zurich. Coding issues and the definition of the variables gathered are discussed in Appendix B.

The vocational school grades collected for this study stem from the first and third semester of vocational school. Thus, since the apprenticeship as retail sales specialist lasts six semesters, the grades considered here represent training success in an early phase of the apprenticeship. However, this is no major drawback. Firstly, grades from the first semesters are relatively closely linked to the grades received at the final exam at the end of the apprenticeship. Secondly, the use of early grades ensures a higher external validity of the study since the specific firm cannot yet strongly influence them.

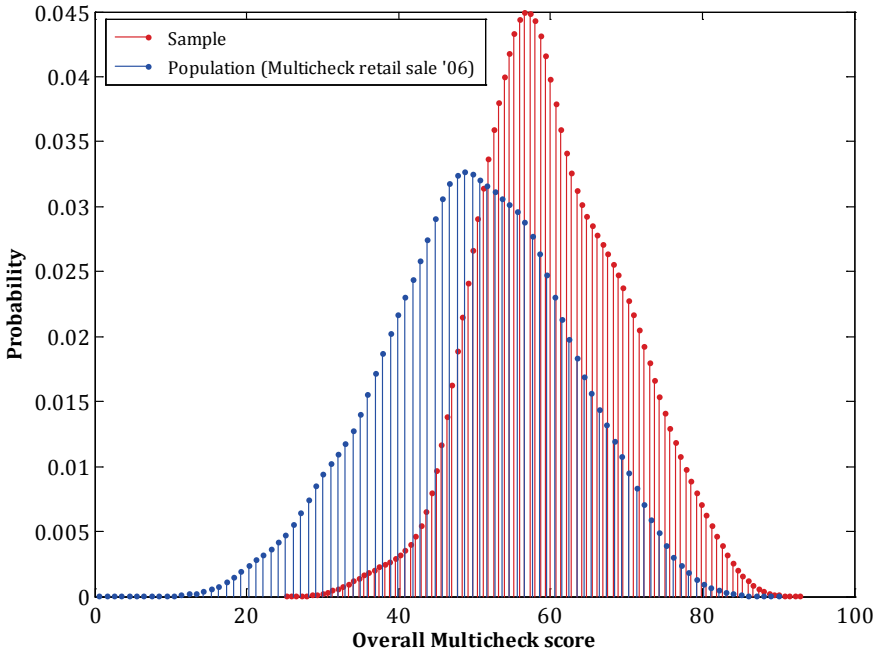
Another mentionable fact contributing to a greater external validity of the study is that vocational training and learning situations of apprentices vary across and within each cooperative. Firstly, hiring and management of apprentices differ between the three

cooperatives, and secondly, apprentices are to a great extent trained on-the-job by an apprenticeship trainer of the branch offices where they work.

However, the study suffers partly from a selection problem: the sample has too few "insufficient" overall Multicheck scores. In particular, the sample probability that an applicant with an insufficient test score is employed at one of the three Migros cooperatives is only 1.2 percent. The pattern arises because the Migros cooperatives put considerable weight on the Multicheck score in the selection of apprentices, and because adolescents with low test scores may stop searching for an apprenticeship as retail sales specialist. Figure 1 illustrates the problem in comparing the sample distribution of the Multicheck overall score (*Gesamtresultat*) with the distribution of overall scores from the population of test-takers (using a Gaussian kernel for illustration purpose).[10] As a result of the compressed distribution of Multicheck overall score, we can only examine the informational value of the upper two thirds of the score. I will discuss the implications of the sample selection problem in the robustness section.

[10] The latter data was provided by Multicheck. The firm sent raw data of the test results of all test takers that passed the Multicheck 06/07 (version of 2006 of the test). Multicheck did not provide the actual overall scores, but only the results in each individual test section. I built an overall test score from the results of the subsections by myself under the condition that the overall score has to have a mean of 50 points according to its scaling (cf. section 2). The resulting figure was approved by Multicheck.

Fig. 1: Sample distribution of overall Multicheck scores (red) and (estimated) population distribution of overall scores (blue)



4. Results

Table 1 shows cross-correlations of the five most important variables. Evidently, the overall test score of the Multicheck does not (or even negatively) correlate with school grades from lower secondary school. This is consistent with the findings of Widmer (2006) for the "Multicheck commercial sector." The correlations between lower secondary school grades and GPA in vocational school lie within the expectations.

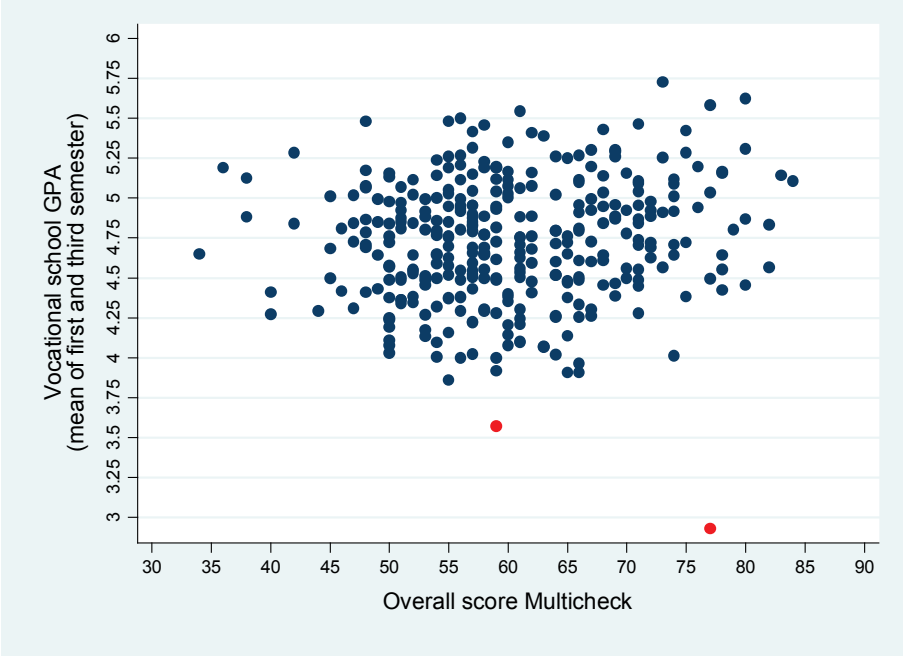
However, the "predictive validities" of 0.113 and 0.12 of the Multicheck overall score do not meet the expectations. Not only are they relatively low compared to the correlations reported in Widmer (2006), but they are also substantially below the correlation of 0.4 established in the Multicheck (2010) study that considers the Multicheck retail sale as well. In addition, the predictive validity of the overall score is worse than those commonly established for tests of general cognitive ability. These lie between 0.3 and 0.5 (see e.g. Schmidt and Hunter 1998).

Table 1: Correlations between lower secondary school GPA, Multicheck overall score and vocational school GPA

	(1)	(2)	(3)	(4)	(5)
GPA of newest lower secondary school record (1)	**1** n=334				
GPA of second newest lower secondary school record (2)	**0.812**** n=326	**1** n=326			
Overall score Multicheck (3)	**-0.011** n=334	**-0.094*** n=326	**1** n=334		
GPA in vocational school, first semester (4)	**0.2747**** n=331	**0.22**** n=323	**0.113**** n=331	**1** n=331	
GPA in vocational school, third semester (5)	**0.291**** n=324	**0.223**** n=316	**0.119**** n=324	**0.737**** n=323	**1** n=324
** p<0.01, * p<0.05					

The small correlation between Multicheck overall score and the average vocational school grade is illustrated in Figure 2. Note that the grading scale in Switzerland ranges from the "excellent" grade 6.0 to 1.0, with 4.0 as the lowest sufficient grade. The scatterplot shows a considerable dispersion of values. It demonstrates that an overall Multicheck result which "exceeds" the requirements, i.e. a score above 60 points, is not a secure sign of good academic performance in vocational school. Similarly, some apprentices with relatively bad Multicheck results (between 40 and 60 points in the overall score) performed well in vocational school. When looking at the figure, one is left with the impression that the Multicheck cannot assure that adolescents with an "insufficient" test score (i.e. Multicheck results with less than 40 points) will effectively underperform in vocational school.

Fig. 2: Overall score in the Multicheck test and average grade in vocational school



Though, as discussed in section 2, this paper aims at establishing the predictive power of the Multicheck in a multivariate regression model. Scrutiny of the data reveals that the outcome variable–vocational school grades–shows quite strong clustering. In particular, the levels and, to a lesser extent, the variances of vocational school grades differ statistically significantly across the 23 vocational schools of the sample despite equal performance standards for all vocational schools. An elegant way of dealing with the problem of clustering in the outcome variable is to use a nonlinear probability model with an appropriate outcome variable (cf. Fielding et al. 2003). Therefore, we construct a grade variable that ranks an apprentice *relative* to the performance of apprentices attending the same vocational school. More precisely, the new outcome variable (named $Grader$) takes on the value zero (low grader) if the average grade of an apprentice belongs to the lowest quartile of the sample distribution of grades in the vocational school he goes to. It is two if he belongs to the quartile of top graders of his school. Else, it is one (medium grader).

One might say that the procedure amounts to assuming that the observed differences in vocational school grades are only due to different degree of difficulties in the vocational

schools or, in other words, to assuming that the apprentices in different vocational schools are equally able. Firstly, this would be not very restrictive, as the selection into vocational schools is largely exogenous and the selection of Migros should make sure relative homogeneity of the sample.[11] And secondly, the variable is a meaningful operationalisation of success in vocational training even if this exogeneity is not given: it is valuable information for the employer to know whether an apprentice is likely to belong to the group of top- or to the group of underperformers in a specific vocational school.[12]

The transformation of vocational school GPA in a categorical variable has a further favourable effect in our context: we only assume that vocational school grades are hierarchical, but not that they are defined by a continuous interval scale (as is necessary in a linear regression model with GPA as a dependent variable). The lack of an interval scale underlying school grades is the reason why several authors have argued that grades should be analysed using an ordered probit or ordered logit model (Fielding et al. 2003; Grant 2007; Sund 2009).[13]

To identify our model, we need to specify a latent variable. Specifically, we have to estimate two threshold values (α_1 and α_2) since the categorical variable has three different realisations:

$$Grader_i = \begin{cases} 0 \; if \; -\infty < Grader_i^* < \alpha_1 \\ 1 \; if \; \alpha_1 < Grader_i^* < \alpha_2 \\ 2 \; if \; \alpha_2 < Grader_i^* < \infty \end{cases} \quad (1)$$

[11] The selection into different vocational schools is determined by the apprentice's branch office. As such, it is exogenous to the apprentice. And since the recruiters at Migros were not aware of different grading practices at vocational schools, and since the placement into branch offices is largely determined by the domicile of the apprentice and job openings at branch offices, it seems exogenous to the firm as well.

[12] One might nevertheless argue that such an outcome variable is a theoretical construct with no direct practical relevance. Another reason justifies the approach chosen here, though: trainers of apprentices in firms are likely to at least partially know the differences in grading between vocational schools. They will thus be able to project the relative outcome variable considered here on an approximate grading scale.

[13] Although the assumption that grades are interval-scaled might not be very strong for the Swiss grading scale, it is quite surely inappropriate in Anglo-Saxon countries in which grades are letters rather than figures.

In this formula, $Grader_i^*$ denotes the underlying latent variable. We estimate the following model for it:

$$Grader_i^* = \beta_1 + \beta_2 MCoverall_i + \beta_3 GPA_i + \boldsymbol{\beta}\mathbf{Z_i} + \varepsilon_i \qquad (2)$$

In (2), $MCoverall_i$ stands for the overall score in the Multicheck retail sale, GPA_i for the average school grade of the two lower secondary school records collected, $\mathbf{Z_i}$ is a vector of control variables, and ε_i is the residual of the regression. The vector of exogenous control variables, $\mathbf{Z_i}$, comprises, in a smaller model, (dummy) variables for age, gender and native language, a variable whether an applicant comes from a lower secondary school in an urban or rural region, variables representing the different types of schools and a dummy whether an apprentice attended a course to bridge gaps in training (CBGT). A larger model adds 11 cantonal dummy variables (with the canton Aargau as reference group) and two dummy variables for the Migros cooperatives (with Migros Aare as reference group) to the vector of controls (cf. Appendix B for the specification of these controls). Note that it is the aim of this paper to estimate an unbiased coefficient for the Multicheck result. This is why some control variables are included in the regression although they are insignificant.

Identification of the model in (2) is achieved by assuming a distribution for the residual ε_i. The most popular choices are the logistic and the standard normal distribution, leading to the ordered logit or ordered probit model, respectively. Test regressions showed that the differences between the two models are minor. I present the results for the ordered probit model.[14]

[14] The main reason is that the fixed-effects model (cf. Appendix) yielded normally distributed residuals, indicating that standard normality may not be an erroneous assumption for the latent variable formulation given in (2).

The ordered probit estimates of equation (2) are given in Table 2.[15] The overall score in the Multicheck test is very inconsistently, if at all, useful as a predictor of average vocational school grades in the first and third semester. The average lower secondary school grade, on the other hand, is a valid predictor of theoretical training success. Moreover, the signalling value (Schwerdt and Wössmann 2009) of secondary school *records* profits from the predictive power of the type of school coefficients. Apprentices that went to school in an intermediate or advanced level school perform significantly better than those from the basic level schools.

Table 2: Determinants of vocational school GPA in the first and third semester (ordered probit regression)

Dependent variable:	*Vocational school GPA 1 (Grader* 1)*		*Vocational school GPA 3 (Grader* 3)*	
	(1)	(2)	(3)	(4)
Overall score MC	0.0091	0.005	0.0133+	0.0102
	(0.007)	(0.008)	(0.007)	(0.008)
GPA	0.644**	0.991**	0.487**	0.730**
	(0.158)	(0.183)	(0.158)	(0.179)
High age	0.169	0.124	-0.126	-0.161
	(0.184)	(0.189)	(0.186)	(0.190)
Low age	-0.079	-0.092	-0.043	-3.88e-05
	(0.257)	(0.267)	(0.257)	(0.267)
Female	-0.0203	-0.0583	0.0863	0.0917
	(0.135)	(0.143)	(0.136)	(0.143)
Language	0.298*	0.389*	0.164	0.209
	(0.147)	(0.153)	(0.150)	(0.154)
Urban	-0.0182	-0.189	-0.064	-0.164
	(0.165)	(0.173)	(0.166)	(0.174)
CBGT	0.097	0.129	0.071	0.086
	(0.148)	(0.154)	(0.148)	(0.154)
Intermediate	0.558**	0.961**	0.410**	0.744**
	(0.143)	(0.174)	(0.143)	(0.172)
Advanced	1.235**	2.014**	1.564**	2.140**
	(0.451)	(0.500)	(0.480)	(0.512)
Cantonal dummies	No	Yes	No	Yes
Dummies for cooperatives	No	Yes	No	Yes
Threshold 1	3.172**	5.119**	2.573**	3.863**
Threshold 2	4.741**	6.787**	4.212**	5.572**
Observations	329	329	324	324
McFadden's Pseudo-R^2	0.069	0.112	0.056	0.088

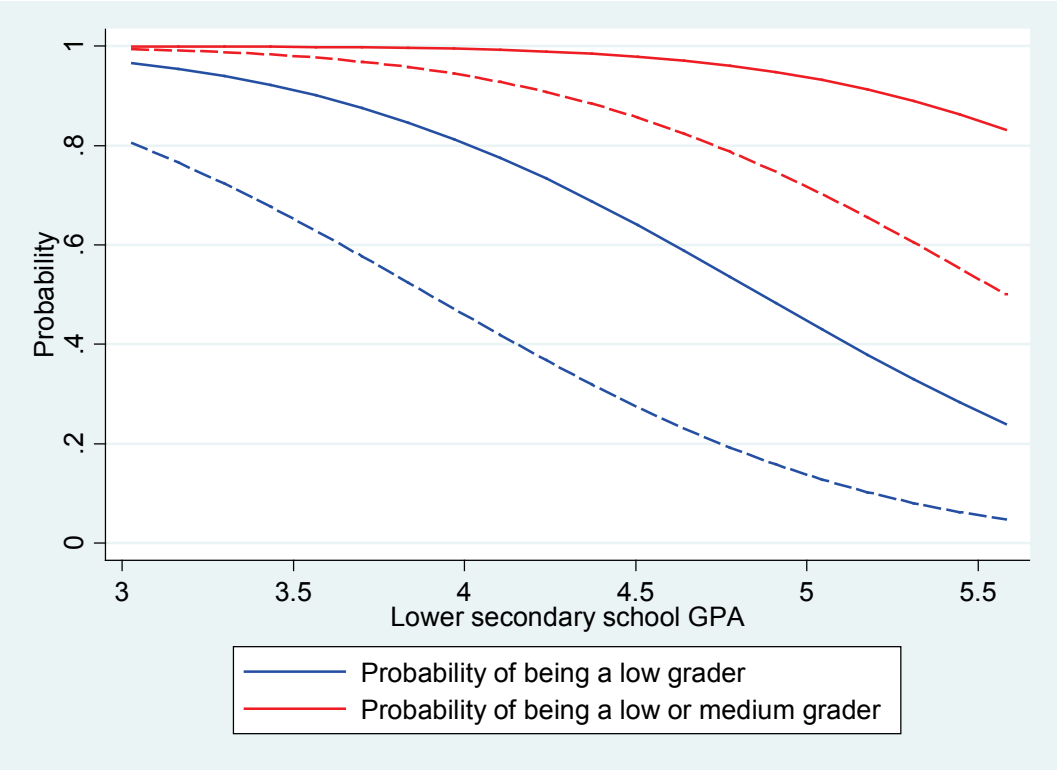
** p<0.01, * p<0.05, + p<0.1

Notes: Standard errors in parentheses.

[15] Two outliers were consequently omitted from the regressions of first semester vocational school GPA because they had a substantial leverage particularly on the coefficient of the Multicheck overall score. The leverage is illustrated in Figure 2, in which the two outliers are highlighted using red dots.

Figure 3 exemplifies the effects of the schooling variables. It shows the probabilities that a hypothetical person, belonging to the reference group[16] and applying with a Multicheck score of 60 points, will be a low grader (straight blue line) or a low or medium grader (straight red line). It also contains the probabilities for a person who did not attend a basic level school (as the person from the reference group), but who attended an intermediate level school instead (dashed lines), everything else held constant. So, for instance, the probability that an adolescent applying with a GPA of 5.5 from an intermediate level school will be a low grader in the first semester of vocational school is below 10 percent.

Fig. 3: Estimated probabilities that a person from the reference group (straight lines) or from an intermediate level school (dashed lines) belongs to the group of low graders or to the group of low and medium graders



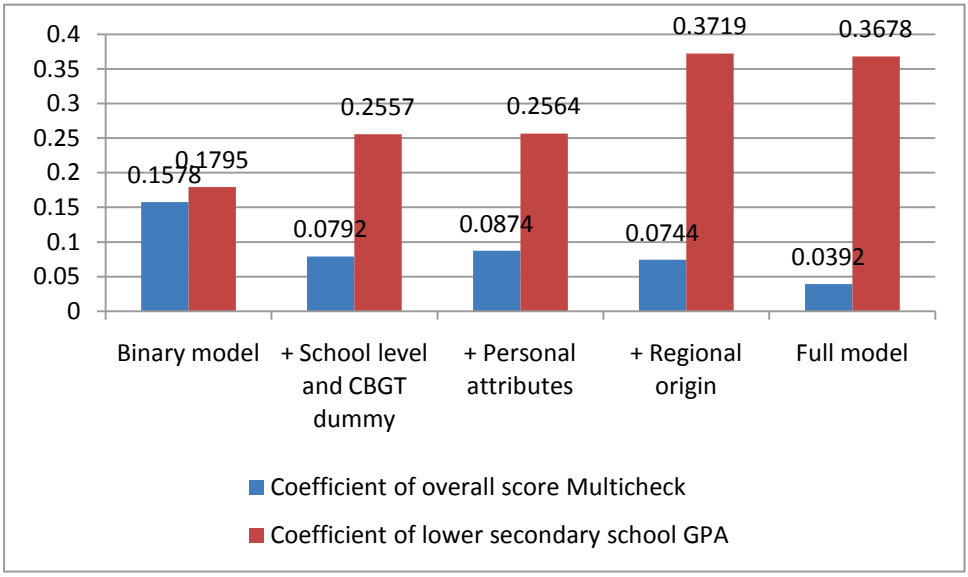
[16] The reference group of the regression of the larger model (models 2 and 4) are male and medium-aged apprentices of Migros Aare that applied from a (rural) basic level school of the canton Aargau, which do not speak German as a native language and which did not attend a CBGT.

An interesting exercise is to examine whether the small informational value of the Multicheck overall score is a result of the inclusion of other variables. Exemplifying a pattern that shows up in all estimated models, Figure 4 provides the answer to this question with the ordered probit model for first semester vocational school GPA. It compares the (standardized) coefficients of the Multicheck overall score with those of the average lower secondary school grade when other explanatory variables are added to the model, beginning with a binary model. The figure illustrates the appropriateness of the multivariate approach. It shows how univariate measures may be misleading concerning the informational surplus the test adds to the selection decision.

Particularly, the overall score of the Multicheck and the GPA have nearly the same informational value to the employer if they are considered independently of any other information. However, the Multicheck looses much of its informational value once the employer knows whether the Multicheck result is from an attendant of a basic, intermediate or advanced level school or from an adolescent who attended a CBGT (second column). This is because attendants from higher school levels or from a CBGT have higher Multicheck scores (as OLS regressions of the Multicheck overall score show). Thus, the Multicheck mirrors the higher average skills of pupils of higher school types, but the signal "type of school" is the stronger predictor of vocational school grades than the Multicheck. The predictive power of the test reduces further if we include lower secondary school grades into the model (last column). In conclusion, the informational surplus the test adds to the selection of retail sales specialists gets minor once the recruiter considers the lower secondary school record of the applicant. This finding is contrary to the hypothesis formulated in section 2.

Interestingly, the pattern is reversed if we look at lower secondary school GPA: the signalling value of school grades increases with the amount of information about their context. In particular, employers can interpret school grades more appropriately if they consider school level and regional origin of the adolescent.

Fig. 4: Standardized coefficients of lower secondary school GPA and the Multicheck overall score when covariates are added to an ordered probit regression



5. Robustness

The Multicheck is a poor predictor of training success of apprentices at Migros. This is a robust result. Firstly, it appears in all specifications of the ordered probit model. Secondly, the result is confirmed in a school-fixed-effects model of vocational school grades, that is, in a linear regression model that corrects for the differences in mean grades across vocational schools (cf. Appendix C). Additionally, the result is insensitive to the inclusion of the variable of social background–a variable of the skill of the apprentice's parents (cf. Appendix B).

However, two qualifications of the finding have to be made. The first one can be seen in Figure 2: the correlation between Multicheck overall score and vocational school grades is greater for Multicheck results above the sample mean (60.17 points) than below it ($\rho_{MCoverall\leq 60}$ = 0.052 and $\rho_{MCoverall>60}$ = 0.281). The result is confirmed in ordered probit regressions, too: the above-average Multicheck results are slightly more valid in predicting vocational school grades than below-average scores.

The non-constancy of the Multicheck coefficient indicates a limitation of this study since it is probably related to the sample selection problem discussed in section 3. It is possible that we lack bad graders for small Multicheck results (observations in the lower-left part in Figure 2) relatively more than we lack them for higher Multicheck scores. For example, adolescents who received an apprenticeship although they only had 40 points in the Multicheck might have written, in order to compensate the bad Multicheck score, a better application letter than those with the same Multicheck result who were not recruited. If the necessity to compensate a bad Multicheck score decreases with a better test result, the unobservable attribute would be (negatively) correlated with the Multicheck score in our sample. Finally, since the unobservable characteristic signalled by the letter is (positively) correlated with vocational school grades, the estimate of the Multicheck coefficient would be biased (downwards).

If such selection biases occurred, then we would not be able to judge the "usefulness" of the Multicheck in the selection of Migros, i.e. its value as a predictor of training success of *applicants*. We could just draw conclusion about the test's predictive power of training success of actual *apprentices*.

The quantitative importance of the selection effect on the Multicheck coefficient is difficult to assess.[17] An attempt can be made by looking at the size of the Multicheck coefficient for high test results. Since the importance of the Multicheck result in the pre-selection gets minor once the candidates "exceed" the requirements of an apprenticeship, it is unlikely that unobservable attributes are strongly correlated with high test scores. As the point estimate of the Multicheck coefficient is also small for test results above 60 points, we can be rather sure that the influence of sample bias on the Multicheck coefficient is small. In addition, note that lower secondary school grades are also exposed to the selection problem. In comparing the coefficient of school grades with the one of the Multicheck, the relative bias should be rather small.

A second qualification of the results is that several subsections of the Multicheck perform reasonably well in predicting vocational school grades. In particular, the language sections of the test are good predictors of corresponding grades in vocational school. For example, the correlation between the French score and the vocational school grade in French is 0.412. Furthermore, Multicheck scores in English and retentiveness are even better predictors of vocational school GPA than overall test scores. Finally, it seems as though the sections "logic" and "concentration" cannot add to or may even undermine the predictive power of the test. Either the test fails in measuring these cognitive skills, or cognitive skills are not decisive for success in vocational training, as would be consistent with the findings of Stamm et al. (2010) mentioned earlier. These results provide strong evidence that the Multicheck suffers from conceptual problems.

[17] The estimation of a selection model that incorporates and models the selection decision of the firm was not feasible (due to privacy concerns and time constraints).

One such problem could be that Multicheck does not sufficiently adjust the content of the test to the specific requirements of the apprenticeship (although the firm claims to do that on its homepage). Subjects and contents important for retail sales specialists in vocational school but not taught in lower secondary school are not covered in the test. As a result, the Multicheck overall score does not reliably predict grades in the vocational school subjects economy, society, general aspects of the branch and retail sales skills.

6. Other variables of success in vocational training

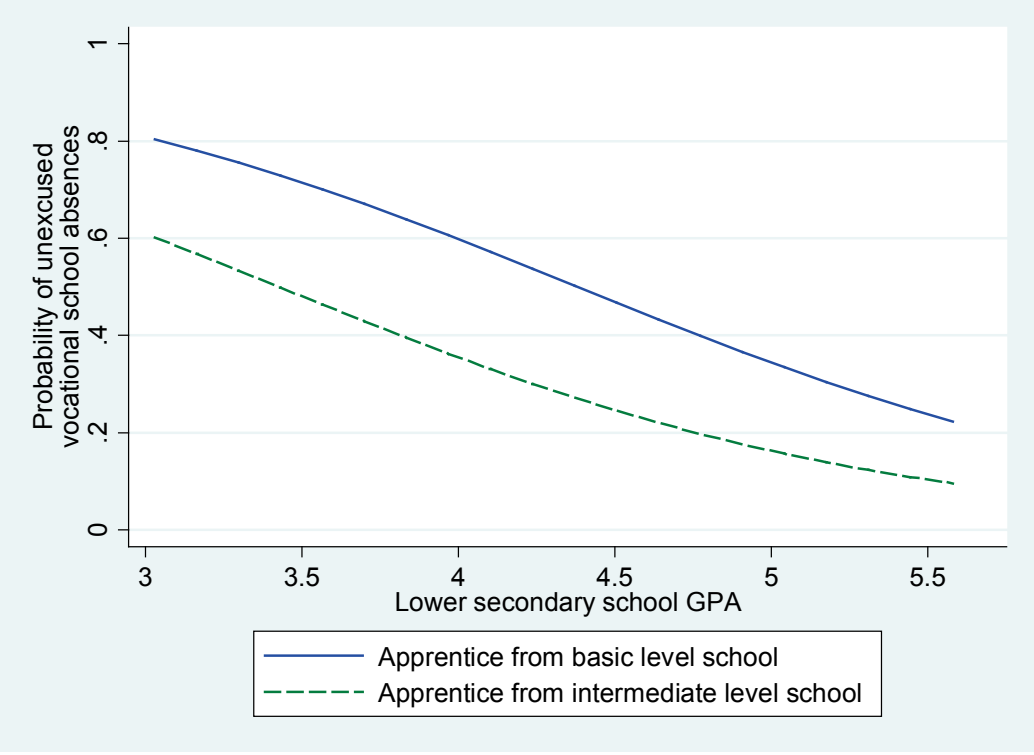
Even if vocational school grades are the most important determinant of theoretical success in vocational training, they only partially reflect the aptitude of an adolescent to successfully pass the apprenticeship as retail sales specialist. Is the Multicheck a better predictor of other criteria of vocational success? Two further outcome variables are analysed in order to answer this question.

The first one is the probability of unexcused vocational school absences which are used as a proxy for the social competences of an apprentice. Unexcused absences in vocational school often express lack of motivation, problems in school, or a missing sense of duty (Lounsbury et al. 2004). Not surprisingly, they are a major reason of premature terminations of apprenticeship contracts (Moser et al. 2008).

Since non-attendance can be the result of problems in school, one might expect that school grades as well as the Multicheck can act as a significant predictor of this outcome variable. The binary variable has a value of one if an apprentice missed school at least once without excuse in the first two years of the apprenticeship. This was the case for 16.55 percent or 48 out of 290 apprentices of the sample (cf. Appendix B for a discussion of the reasons for the drop in sample size).

The results of logit regressions of the outcome variable on the set of control variables already used in the models of section 4 are presented in the Appendix.[18] Apparently, lower secondary school GPA adds to the prediction of unexcused absences in vocational school, and is even the only valid predictor in the smaller model. In the larger model, the coefficient of type of school is also significant. Figure 5 illustrates the predictive power of these two variables. An applicant who belongs to the reference group of the regression applying with a GPA of 4.0 has a considerably higher probability to cause unexcused vocational school absences than an applicant from an intermediate level school with a GPA of 5.0, ceteris paribus. The Multicheck overall score, on the other hand, has no predictive power concerning unexcused absences of apprentices.

Fig. 5: Probability of unexcused vocational school absences according to lower secondary school GPA and type of school



[18] The logit model was chosen as it slightly outperforms the probit counterpart concerning the information criteria. One might fear that the logit regressions suffer from clustering. This is reasonable since the probability of having unexcused absences significantly differs across vocational school. Therefore, several logit and probit regressions–school-fixed-effects logit, and random-effects probit–were run in order to control the results. The main coefficients are nearly identical across models. Furthermore, the standard deviations of the main coefficients are even smaller than those given in Table 4 if we use a cluster-robust estimator.

It is not negligible that the Multicheck fails to predict unexcused absences in vocational school: disciplinary reasons are the main determinant of premature apprenticeship terminations in the Migros (according to the trainers of apprentices)–which is, arguably, the most evident sign of failure in vocational training. Can the test nonetheless predict a premature apprenticeship termination? Note that this would be desirable considering the verdict the test returns on an adolescent's "ability" for a certain apprenticeship.

It was, unfortunately, not possible to collect the data for all of apprentices that started their apprenticeship mid-2007 or 2008 but dropped out. As a result, the data set contains only 14 drop-outs in a sample of 250 apprentices (see Appendix B for a discussion of this issue). Nevertheless, the binary variable indicating a drop-out was regressed on the Multicheck overall score, the lower secondary school GPA, and some control variables.[19]

The results of the probit regression are shown in Table 3. These results are not fully reliable since they depend relatively strongly individual drop-outs and since they base upon an incomplete sample. However, two results were robust in all specifications, subsamples and models tested: while the Multicheck overall score is not able to predict the probability of a premature apprenticeship termination, GPA from lower secondary school does, at least on a significance level of 10 percent. The latter finding is notable because only four of the 14 drop-outs were explicitly due to problems of meeting academic requirements in vocational school. All other premature apprenticeship terminations were due to disciplinary problems, insufficient practical performances or wrong choice of occupation.

[19] Some explanatory variables of the regressions above were excluded as they were insignificant and did not influence the results of the main coefficients. An inclusion of all covariates used in the models above did not seem meaningful as it should be estimated efficiently having so few observations.

Table 3: Determinants of premature apprenticeship terminations (Probit regression)

Dependent variable:	*Premature apprenticeship termination*
	(1) Probit
Overall score MC	0.021
	(0.0173)
GPA	-0.632+
	(0.332)
Female	0.325
	(0.301)
Language	0.191
	(0.311)
Urban	0.315
	(0.325)
Intermediate	-1.148**
	(0.443)
Constant	0.281
Observations	250
McFadden's Pseudo-R^2	0.116

** p<0.01, * p<0.05, + p<0.1
Notes: Standard errors in parantheses

7. Concluding remarks

Lower secondary school GPA and the type of school the applicant attended are a reliable source of information for a firm searching an apprentice. The signalling value of school grades in this study is remarkable, as we know from earlier work that school grades tend to be systematically biased and noisy, and since grades signal scholastic or cognitive abilities only very imperfectly. Why could school grades nevertheless be valid predictors of vocational training success? It might be the joint result of the reduction in the noise and bias in grades when they are averaged (cf. Grant 2007), and of the positive relation between school grades and non-cognitive skills of pupils, which in turn positively influence training success.[20]

[20] Studies have shown that grades contain a social component, depending on the character (above all frankness, friendliness and sense of duty), interest and motivation of the pupil (Baron-Boldt 1989; Imdorf 2007; Lekholm and Cliffordson 2009; Miller 1998).

On the other hand, the Multicheck retail sale cannot meet the expectations firms have in it. The test is not a valid predictor of training success of apprentices at Migros. It seems as though a substantial part of the small but positive correlation between overall Multicheck score and vocational school grades stems from the fact that the test result is better for apprentices who attended a higher school level–an information that the firm can infer from the application dossier. In addition, the test does not add to the prediction of inappropriate social behaviour (the probability of unexcused absences in vocational school) or of the likelihood of a premature apprenticeship termination. If one is willing to believe that these results do not strongly suffer from a sample selection bias, we may conclude that the test does not significantly improve Migros' selection of apprentices.

Yet, do the findings carry over to other circumstances, apprentices and firms as those examined in this investigation, that is, are they externally valid? In this context, it is important to note that the results contain a substantial amount of variation in individual training and learning situations and are hence valid in very different settings. The apprentices not only come from twelve different cantons, dozens of different lower secondary schools with different types of schools, but they are also educated in 23 different vocational schools, in three different cooperatives and in nearly a hundred individual branch offices. Nevertheless, we cannot be certain that the results are generalisable to other firms, particularly as studies have shown that vocational success is to a considerable extent firm-specific (e.g. Stalder and Schmid 2006; Stamm et al. 2010) and that firms differ with respect to their recruitment of apprentices (Imdorf 2009; Moser 2004). Thus, the results of this study might be specific to the context of the Migros cooperatives examined.[21]

[21] They can definitely not be generalised to other versions of the Multicheck test since, for example, the conceptual problems of the test might not carry over to them.

In any case, even if the results were only valid for the Migros, they would be important: Migros represents more than one third of the Swiss retail business. And the study yields another result which is problematic for the test: the weak predictive validity (as, e.g., illustrated in Figure 2) and the conceptual problems mentioned strongly suggest that the test fails to measure the "aptitude" of test takers as retail sales specialists. This conclusion can be drawn irrespective of possible sample selection or firm effects: as the test aims to indicate whether an adolescent is able or unable for the apprenticeship as retail sales specialists, it should be diagnostically correct regardless of the specific context of the firm or of unobserved attributes (or else it should measure them). In conclusion, it seems as though the test is more an assessment of knowledge taught at school than a proper "aptitude" test for apprenticeships.

References

AMOS, J., AMSLER, F., MARTIN-JAHNCKE, M., and MICHEL, B., (2009). ***Evaluation der Resultate von Lehrabschlussprüfungen der beruflichen Grundausbildungen*****. Büro für Kommunikation, Basel.**

ARROW, K. J., (1973). "Higher education as a filter." ***Journal of Public Economics*****, 2(3), 193-216.**

BARON-BOLDT, J., (1989). ***Die Validität von Schulabschlussnoten für die Prognose von Ausbildungs- und Studienerfolg: eine Metaanalyse nach dem Prinzip der Validitätsgeneralisierung*****, Frankfurt a. M.: Verlag Peter Lang.**

BAUER, P., and SHELDON, G., (2008). ***Ethnic Discrimination in Education: The Swiss Case*****, Basel: Department of Economics, FAI.**

BERTSCHY, K., BÖNI, E., and MEYER, T., (2008). ***Young people in transition from education to labour market. Results of the Swiss youth panel survey TREE, update 2007*****. TREE, Basel.**

BFS, (2005). ***PISA 2003: Kompetenzen für die Zukunft. Zweiter nationaler Bericht*****. BFS, Neuenburg.**

BISHOP, J. H., (1994). "Signaling the competencies of high school students to employers." ***CAHRS Working Paper 94-18*****.**

COSTRELL, R. M., (1994). "A simple model of educational standards." ***The American Economic Review*****, 84(4), 956-971.**

DE PAOLA, M., and SCOPPA, V., (2010). "A signalling model of school grades under different evaluation systems." ***Journal of Economics*****, 101(1), 1-14.**

FIELDING, A., YANG, M., and GOLDSTEIN, H., (2003). "Multilevel ordinal models for examination grades." ***Statistical Modelling*****, 3(2), 127-153.**

GRANT, D., (2007). "Grades as information." ***Economics of Education Review*****, 26(2), 201-214.**

HÄBERLIN, U., IMDORF, C., and KRONIG, W., (2004). ***Von der Schule in die Berufsschule. Untersuchungen zur Benachteiligung von ausländischen und von weiblichen Jugendlichen bei der Lehrstellensuche*****, Bern, Stuttgart, Wien: Haupt.**

HECKMAN, J., and VYTLACIL, E., (2001). "Identifying the role of cognitive ability in explaining the level of and change in the return of schooling." ***The Review of Economics and Statistics*****, 83(1), 1-12.**

HUNSLEY, J., and MEYER, G. J., (2003). "The Incremental Validity of Psychological Testing and Assessment: Conceptual, Methodological, and Statistical Issues." ***Psychological Assessment*****, 15(4), 446-455.**

IMDORF, C., (2005). ***Schulqualifikation und Berufsfindung. Wie Geschlecht und nationale Herkunft den Übergang in die Berufsbildung strukturieren*****, Wiesbaden: Verlag für Sozialwissenschaften.**

IMDORF, C., (2007). "Die relative Bedeutsamkeit von Schulqualifikationen bei der Lehrstellenvergabe in kleineren Betrieben", in T. Eckert, (Hrsg.), ***Übergänge im Bildungswesen*****. Münster: Waxmann Verlag.**

IMDORF, C., (2009). "Die betriebliche Verwertung von Schulzeugnissen bei der Ausbildungsstellenvergabe." ***Empirische Pädagogik*****, 23(4), 392-409.**

KRONIG, W., (2007). ***Die systematische Zufälligkeit des Bildungserfolgs*****, Bern, Stuttgart, Wien: Haupt.**

LEKHOLM, A. K., and CLIFFORDSON, C., (2009). "Effects of student characteristics on grades in compulsory school." ***Educational Research and Evaluation: An International Journal on Theory and Practice*****, 15(1), 1-23.**

LINDAHL, E., (2007). ***Comparing teachers' assessments and national test results – evidence from Sweden*****. IFAU - Institute for Labour Market Policy Evaluation.**

LOUNSBURY, J. W., STEEL, R. P., LOVELAND, J. M., and GIBSON, L. W., (2004). "An investigation of personality traits in relation to adolescent school absenteeism." ***Journal of Youth and Adolescence*****, 33(5), 457-466.**

MILLER, S. R., (1998). "Shortcut: high school grades as a signal of human capital." ***Educational Evaluation and Policy Analysis*****, 20(4), 299-311.**

MOSER, C., STALDER, B. E., and SCHMID, E., (2008). ***Lehrvertragsauflösungen: Die Situation von ausländischen und Schweizer Lernenden. Ergebnisse aus dem Projekt LEVA*****, Bern: Erziehungsdirektion des Kantons Bern (ERZ).**

MOSER, U., (2004). ***Jugendliche zwischen Schule und Berufsbildung. Eine Evaluation bei Schweizer Grossunternehmen unter Berücksichtigung des internationalen Schulleistungsvergleichs PISA*****, Bern: h.e.p. verlag.**

MÜHLEMANN, S., WOLTER, S., FUHRER, M., and WÜEST, A., (2007). ***Lehrlingsausbildung – ökonomisch betrachtet. Ergebnisse der zweiten Kosten-Nutzen-Studie*****, Zürich, Chur: Rüegger Verlag.**

MÜHLEMANN, S., WOLTER, S. C., and WÜEST, A., (2009). "Apprenticeship training and the business cycle." ***Empirical Research in Vocational Education and Training*****, 1(2), 173-186.**

MULTICHECK, (2010). ***Multicheck – Wissenschaftliche Erhebung*****, http://multicheck.potentials.ch/fileadmin/user_upload/Wissenschaftliche_Dokumente/Broschuere_Wissenschaft.pdf, Konolfingen: Multicheck. Stand: 23.08.2010.**

MURNANE, R. J., WILLETT, J. B., BRAATZ, M. J., and DUHALDEBORDE, Y., (2001). "Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY." ***Economics of Education Review*****, 20(4), 311-320.**

SCHMIDT, F. L., and HUNTER, J. E., (1998). "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings." ***Psychological Bulletin*****, 124(2), 262-274.**

SCHWERDT, G., and WÖSSMANN, L., (2009). "The signaling value of central school exams." ***Preliminary Version*****.**

SPENCE, M., (1973). "Job market signaling." ***The Quarterly Journal of Economics*****, 87(3), 355-374.**

STALDER, B. E., MEYER, T., and HUPKA-BRUNNER, S., (2008). "Leistungsschwach – Bildungsarm? Ergebnisse der TREE-Studie zu den PISA-Kompetenzen als Prädiktoren für Bildungsschancen in der Sekundarstufe II." ***Die Deutsche Schule*****, 100(4), 436-448.**

STALDER, B. E., and SCHMID, E., (2006). ***Lehrvertragsauflösungen, ihre Ursachen und Konsequenzen. Ergebnisse aus dem Projekt LEVA*****. Erziehungsdirektion des Kantons Bern, Bern.**

STAMM, M., NIEDERHAUSER, M., and KOST, J., (2010). "The top performers in vocational training." ***Empirical Research in Vocational Education and Training*****, 2, 65-81.**

STIGLITZ, J. E., (1975). "The theory of "screening", education, and the distribution of income." ***The American Economic Review*****, 65(3), 283-300.**

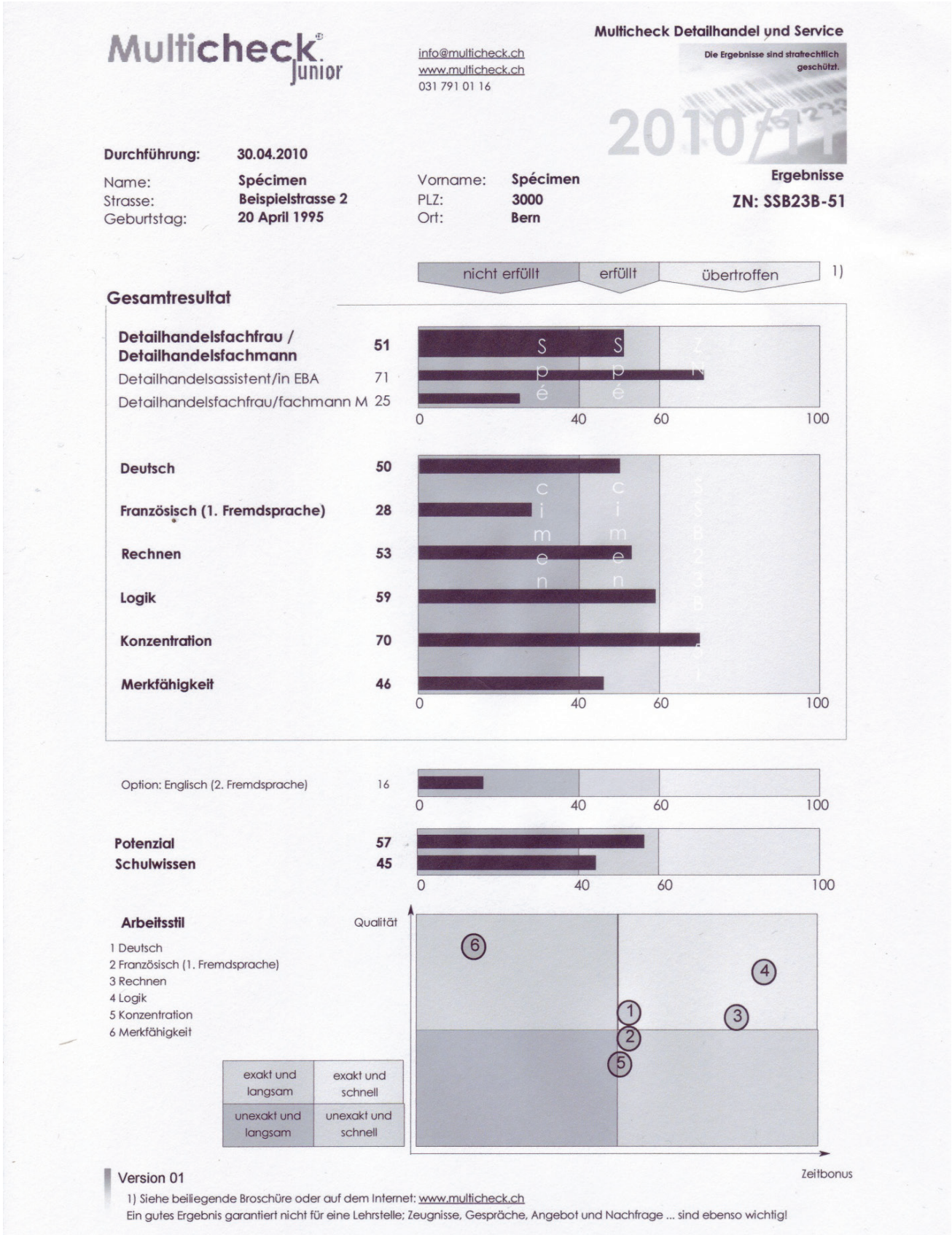
SUND, K., (2009). "Estimating peer effects in Swedish high school using school, teacher, and student fixed effects." ***Economics of Education Review*****, 28(3), 329-336.**

WEISS, A., (1995). "Human capital vs. signalling explanations of wages." ***Journal of Economic Perspectives*****, 9(4), 133-154.**

WIDMER, M., (2006). ***Reliabilität und Validität eines Berufseignungstests bei kaufmännischen Lehrlingen*****, Universität Bern, Bern.**

Appendix

A. Multicheck evaluation sheet



B. Definition of variables and coding

Variables from the CV
• CBGT: dummy variable indicating whether an apprentice passed a transitional year in a course to bridge gaps in training.[22] Here, a CBGT is any employment had or course passed in a transitional year that is likely to have increased scholastic abilities beneficial for vocational school. • Gender • High age, low age: dummy variables indicating whether an apprentice belongs to the highest (older than 17.67) or lowest (younger than 16) quintile of the sample age distribution of apprentices at the start of the apprenticeship. The use of dummy variables instead of a continuous variable is motivated by the presumption of nonlinear age effects. • Language: dummy variable indicating whether the native language of an apprentice is German or not • Skill of parents: dummy variable that acts as proxy for the social background of an apprentice. It bases upon the information about the professions of the parents which is available in most CVs. These professions were coded according to Austrian version of the International Standard Classification of Occupations (Ö-ISCO). The classification groups all professions hierarchically into ten different categories which in turn correspond to four skill levels. The variable bases upon these skill levels. The dummy takes the value one if the profession requires a skill level of three or four. The higher skill level of either of the parents is decisive for the assignment. The variable is not part of the standard set of control variables used in the regressions because it could be coded for only 247 apprentices and had no major impact on the results of the regressions.
Variables from lower secondary school reports
• GPA: Average grade built from the newest and second newest lower secondary school reports. Only grades from the least common denominator of subjects were collected as cantons differ greatly in their curricula. This is, grades from mathematics, German, French, English, History and Geography (taken together) and natural sciences are considered. Grades range from 6.0 (excellent) to 1.0 in Switzerland, commonly in steps of 0.5. 4.0 is the lowest sufficient grade. I always considered grades from the two newest lower secondary school reports irrespective of the fact that these reports may stem from the eighth, the ninth or even the tenth

[22] These courses are offered as transitional solutions for young people who, after completing the lower secondary level, do not immediately begin basic vocational education and training or do not continue their education at a school on the upper secondary level.

school year. Comparing grades across school years is justified because the firm will hire an apprentice based on the newest school records.

- Type of school (basic, intermediate or advanced school level): Types of schools were classified into three different performance levels, similar as in the publications of the Swiss Federal Statistical Office (e.g. BFS 2005). However, this is a simplification. Some cantons in Switzerland place pupils at the end of the sixth (or fourth) school year into two or four different type of schools according to their performance level. In addition, school systems vary also in the amount of permeability of the school levels.
- Urban: dummy variable indicating whether the ancient school of an apprentice was situated in an urban or rural area. The variable was coded using the spatial typology of the Federal Statistical Office. This typology groups regions into four spatial categories: cities (1), agglomerations (2), two categories for rural regions (3 and 4). Consequently, any school situated in an area with an index lower than or equal to two is treated as "urban".

Variables from the Multicheck evaluation sheet

- Overall score MC: Multicheck overall score ranging from 1 to 100
- Test scores of subsections of the Multicheck ranging from 1 to 100: German, French, English, calculating, logic, concentration, retentiveness

Variables from vocational school reports

- Dummy variables for the Migros cooperative that the apprentice works for (Aare, Lucerne, and Zurich)
- Unexcused absences: dummy variable which is one if an apprentice had unexcused absences from vocational school during the first two years of the apprenticeship. The variable is binary instead of continuous because some vocational schools do not provide information about absences in the school records. Trainers of apprentices provided the missing information for the apprentices from Migros Zurich, but just in a binary form. The sample size nevertheless drops to 290 since information about absences was unavailable for some apprentices from Migros Aare and Lucerne.
- Premature apprenticeship termination: dummy variable indicating whether an apprenticeship contract had to be terminated prematurely. Unfortunately, the application material from several drop-outs had already been sent back to the adolescent, making it impossible to collect the data from all drop-outs. Furthermore, Migros Lucerne did not provide information from their drop-outs because of privacy concerns. All apprentices from this cooperative have been excluded from the analysis.
- Vocational school GPA 1 and vocational school GPA 3: mean vocational school grade in the first and third

semester, respectively, basing on grades in all subjects taught at vocational school: economy, a foreign language (either French or English), general aspects of the branch (*Branchenkunde*), German, society, retail sales skills (*Detailhandelskenntnisse*). The latter grade weighs twice in the calculation of the GPA because the subject counts double in the final exam of the apprenticeship.

C. Fixed-effects regression of vocational school grades

The model:

$$VSGPA_{ij} = \beta_1 + \beta_2 MCoverall_{ij} + \beta_3 GPA_{ij} + \boldsymbol{\beta}\mathbf{Z_{ij}} + \zeta_j + \varepsilon_{ij} \quad (3)$$

In (3), $VSGPA_{ij}$ stands for the average vocational school grade (either in the first of third semester), $MCoverall_{ij}$ for the overall score in the Multicheck retail sale, GPA_{ij} for the average school grade of the two school records collected, $\mathbf{Z_{ij}}$ is the vector of control variables explained in the text (section 4), ζ_j represents the school-fixed-effect–nothing different than an intercept specific to each vocational school j–, and ε_{ij} is the residual of the regression.[23]

[23] Robustness checks show that the main results of the FE regression do not hinge upon the inclusion of fixed-effects: OLS and FGLS regressions yielded very similar regression coefficients. Hence, the inclusion of school-fixed-effects as additional covariates does not bias the results concerning the usefulness of the Multicheck and of school grades as screening devices.

Table 4: Determinants of vocational school GPA in the first and third semester (Fixed-effects regressions)

Dependent variable:	*Vocational school GPA 1 (VSGPA 1)*		*Vocational school GPA 3 (VSGPA 3)*	
	(1)	(2)	(3)	(4)
Overall score MC	0.0042*	0.0037+	0.0023	0.0023
	(0.002)	(0.002)	(0.003)	(0.003)
GPA	0.298**	0.293**	0.340**	0.345**
	(0.044)	(0.046)	(0.053)	(0.057)
High age	0.036	0.0327	-0.104+	-0.0945
	(0.0418)	(0.0416)	(0.069)	(0.068)
Low age	0.029	0.0124	0.0168	0.00281
	(0.084)	(0.087)	(0.086)	(0.086)
Female	0.021	0.00836	0.0341	0.0173
	(0.023)	(0.029)	(0.04)	(0.044)
Language	0.140*	0.134*	0.085*	0.078+
	(0.05)	(0.052)	(0.035)	(0.04)
Urban	-0.06	-0.065	-0.142*	-0.141*
	(0.05)	(0.058)	(0.051)	(0.056)
CBGT	0.052	0.0550	0.0224	0.0319
	(0.056)	(0.059)	(0.066)	(0.073)
Intermediate	0.286**	0.281**	0.307**	0.316**
	(0.048)	(0.041)	(0.061)	(0.061)
Advanced	0.486**	0.534**	0.658**	0.733**
	(0.115)	(0.109)	(0.122)	(0.125)
Constant	2.942**	2.845**	2.796**	2.537**
Cantonal dummies	No	Yes	No	Yes
Dummies for cooperatives	No	Yes	No	Yes
Observations	329	329	324	324
R^2	0.263	0.296	0.219	0.256
Cluster	23	23	23	23

** p<0.01, * p<0.05, + p<0.1

Notes: Cluster-robust standard errors in parentheses.

D. Logit regressions of unexcused vocational school absences

Table 5: Determinants of unexcused vocational school absences (Logit regressions)

Dependent variable:	*Unexcused vocational school absences*	
	(1)	(2)
Overall score MC	0.0223	0.0114
	(.0197)	(0.0218)
GPA	-0.9313*	-1.041*
	(0.3941)	(0.460)
High age	0.233	0.304
	(0.446)	(0.478)
Low age	-0.0474	0.218
	(0.6874)	(0.733)
Female	-0.4748	-0.520
	(0.344)	(0.370)
Language	0.0023	0.0229
	(0.389)	(0.409)
Urban	0.455	0.545
	(0.389)	(0.412)
CBGT	0.0425	-0.145
	(0.386)	(0.412)
Intermediate	-0.604	-0.995*
	(0.389)	(0.477)
Advanced	0.6469	-0.160
	(0.9626)	(1.102)
Constant	1.4869	3.877*
Cantonal dummies	No	Yes
Dummies for cooperatives	No	Yes
Observations	290	290
McFadden's Pseudo-R^2	0.0768	0.166

** p<0.01, * p<0.05, + p<0.1

Notes: Standard errors in parentheses.