Universität Bern
Departement Volkswirtschaftslehre
Forschungsstelle für Bildungsökonomie
Prof. Dr. Stefan Wolter

Was kann ein standardisierter Berufseignungstest, was die Schulnoten nicht können?

Eine quantitative Fallstudie zum Signalwert zweier Instrumente der Lehrlingsselektion

Michael Siegenthaler
Master in Economics
05-126-040
Glockenstrasse 9
3018 Bern
Tel.: 079 730 21 70
m.siegenthaler@students.unibe.ch

13. September 2010

Inhaltsverzeichnis

Tabellen- und Abbildungsverzeichnis ... iii
Abbildungsverzeichnis ... iii
Tabellenverzeichnis ... iii
1 Einleitung ... 1
1.1 Problem- und Fragestellung ... 1
1.2 Stand der Forschung ... 3
1.3 Der Multicheck ... 4
2 Die Lehrlingsselektion als Prototyp eines Signalling-Problems ... 6
2.1 Einleitung ... 6
2.2 Theoretische Überlegungen zum Signalgehalt schulischer Leistungsnachweise ... 7
2.2.1 Volksschulnoten als Signal ... 8
2.2.2 Der Schultypus als Signal ... 10
2.3 Der Multicheck als Signal ... 12
2.4 Indizes – unbeeinflussbare Determinanten des Lehrerfolgs ... 14
2.5 „Lehrerfolg“ als Zielgrösse ... 15
2.6 Hypothesen zum Signalwert von Multicheck und Schulnoten ... 16
3 Datengrundlage und Methodik ... 18
3.1 Datengrundlage ... 18
3.1.1 Stichprobe ... 18
3.1.2 Grundgesamtheit und das Problem der Vorselektion ... 20
3.1.3 Datenerhebung und Erfassungsmethodik ... 21
3.2 Variablen und Transformationen ... 23
3.3 Deskriptive Statistik ... 24
3.4 Determinanten des Detailhandels-Multichecks ... 26
4 Vorhersagevaliditäten als Schätzungen des Signalwerts ... 29
4.1 Korrelationen ... 29
4.2 Lineare Regressionsmodelle der Berufsschulnoten ... 31
4.2.1 Modell, Spezifikationen und Ergebnisse ... 31
4.2.2 Robustheit ... 37

4.3 Nichtlineare Wahrscheinlichkeitsmodelle der Berufsschulnoten 39
4.3.1 Problemstellung und Modell 39
4.3.2 Ergebnisse 41
4.3.3 Robustheit 45
4.4 Differenzierungen und Erweiterungen 46
4.4.1 Kollinearität und Signifikanz: kontextuelle Bedingtheit des Signalgehalts 47
4.4.2 Schulnotensignalgehalt nach Schultyp, Kanton und Muttersprache 48
4.4.3 Signalgehalt im Schulzeugnis vermerkter Absenzen 51
4.4.4 Signalwert einzelner Schulnoten und einzelner Multicheck-Testgebiete 52
4.4.5 Modelle einzelner Berufsschulnoten 54
4.5 Andere Kriterien des Lehrerfolgs 56
4.5.1 Unentschuldigte Berufsschulabsenzen 56
4.5.2 Lehrvertragsauflösungen 60
4.6 Externe Validität der Ergebnisse 62
5 Schlussfolgerungen 64
6 Literaturverzeichnis 68
6.1 Literatur 68
6.2 Internetquellen 74
7. Anhang 75
A. Multicheck-Musterauswertung 75
B. Variablen und Codierung 76
C. Histogramme einiger wichtiger Variablen 77
D. OLS- und GLS-Regressionen des Basismodells und des erweiterten Basismodells 80
E. Signalwert nach Multicheck-Durchführungsjahr und Schuljahr des Zeugnisses 81
F. Normalquantil-Plots der Residuen der Fixed-Effects-Regressionen 82
G. Ordered-Probit-Regressionen unter Kontrolle der sozialen Herkunft der Lernenden 83
H. Aufbauende Ordered-Probit-Modelle 84
I. Schulnotensignalwert nach Schultyp 86
J. Fixed-Effects-Modelle der Noten einzelner Berufsschulfächer 87
K. RE- und FE-Regression der Wahrscheinlichkeit unentschuldigter Absenzen in der Berufsschule 94

L. Selbständigkeitserklärung 95

TABELLEN- UND ABBILDUNGSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

ABBILDUNG 1: VERGLEICH DER VERTEILUNG DER IN DER STICHPROBE ERZIELTEN MULTICHECK-GESAMTERGEBNISSE (MCGESAMT, ROT) MIT DER GESCHÄTZTEN VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE DER GRUNDGESAMTHEIT (MULTICHECK VERSION 2006, BLAU) 21
ABBILDUNG 2: STICHPROBENHÄUFIGKEITEN EINZELNER PERSÖNLICHER MERKMALE 23
ABBILDUNG 3: BOXPLOTS DER DURCHSCHNITTLICHEN BERUFSSCHULNOTE IM ERSTEN UND DRITTEN SEMESTER (LSCHNITT13) NACH MIGROS-GENOSSENSCHAFT, GESCHLECHT UND MUTTERSPRACHE 24
ABBILDUNG 4: SIMULIERTE DICHTEFUNKTIONEN DES BERUFSSCHULNOTENDURCHSCHNITTS IM ERSTEN SEMESTER (LSCHNITT1) NACH BESUCHTEM SCHULTYP IN DER VOLKSSCHULE 25
ABBILDUNG 5: TATSÄCHLICHE UND GEFITTETE WERTE EINER BINÄREN REGRESSION DER DURCHSCHNITTSNOTE DER BEIDEN LEHRZEUGNISSE AUF DIE DURCHSCHNITTSNOTE DER BEIDEN BEWERBUNGSZEUGNISSE 30
ABBILDUNG 6: MITTELWERT UND STREUUNG DER BERUFSSCHULNOTEN DES ERSTEN UND DRITTEN SEMESTERS NACH BERUFSSCHULEN (N=23) 32
ABBILDUNG 7: GEMEINSAME VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE (MCGESAMT) UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM ERSTEN SEMESTER 36
ABBILDUNG 8: GEMEINSAME VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE (MCGESAMT) UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM DRITTEN SEMESTER 37
ABBILDUNG 9: MIT DEM ORDERED-PROBIT-MODELL ÜBER DIE SCHULNOTENSKALA GESCHÄTZTE KUMULATIVE WAHRSCHEINLICHKEITEN IM ERSTEN SEMESTER DEM ERSTEN ODER DEN MITTLEREN QUARTILEN DER NOTENVERTEILUNG DER JEWEILS BESUCHTEN BERUFSSCHULE ZUZUGEHÖREN 44
ABBILDUNG 10: GRÖSSE DER STANDARDISIERTEN KOEFFIZIENTEN VON MCGESAMT UND ZSCHNITT12 IM ORDERED-PROBIT-MODELL BEI HINZUNAHME VON KONTROLLVARIABLEN 47
ABBILDUNG 11: TATSÄCHLICHE UND GEFITTETE WERTE EINER BINÄREN REGRESSION DER DURCHSCHNITTLICHEN FRANZÖSISCHNOTEN AUF DAS FRANZÖSISCHERGEBNIS IM MULTICHECK 55
ABBILDUNG 12: MIT DEM LOGIT-MODELL GESCHÄTZTE AUFTRETENSWAHRSCHEINLICHKEITEN UNENTSCHULDIGTER BERUFSSCHULABSENZEN IN DEN ERSTEN DREI LEHRSEMESTERN (LUNENTSCH01) NACH SCHULNOTEN, SCHULTYP UND RÄUMLICHER HERKUNFT 59
ABBILDUNG 13: HISTOGRAMM DER DURCHSCHNITTSNOTE DES LETZTEN BEWERBUNGSZEUGNISSES 77
ABBILDUNG 14: HISTOGRAMM DER DURCHSCHNITTSNOTE DES ZWEITLETZTEN BEWERBUNGSZEUGNISSES 77
ABBILDUNG 15: HISTOGRAMM DES MULTICHECK-GESAMTERGEBNISSES 78
ABBILDUNG 16: HISTOGRAMM DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM ERSTEN LEHRSEMESTER 78
ABBILDUNG 17: HISTOGRAMM DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM DRITTEN LEHRSEMESTER 79
ABBILDUNG 18: NORMALQUANTIL-PLOT DER RESIDUEN DES FIXED-EFFECTS-MODELLS FÜR LSCHNITT1 (ERWEITERTES BASISMODELL) 82
ABBILDUNG 19: NORMALQUANTIL-PLOT DER RESIDUEN DES FIXED-EFFECTS-MODELLS FÜR LSCHNITT3 (ERWEITERTES BASISMODELL) 82

TABELLENVERZEICHNIS

TABELLE 1: LINEARE REGRESSIONSMODELLE DES MULTICHECK-GESAMTERGEBNISSES 27
TABELLE 2: KREUZKORRELATIONEN DER DURCHSCHNITTLICHEN VOLKSSCHULNOTEN, DES MULTICHECK-GESAMTERGEBNISSES UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN 29

Tabelle 3: Kreuzkorrelationen der Multicheck-Testgebiete mit den Durchschnittsnoten der Lehrzeugnisse im ersten und dritten Semester 31
Tabelle 4: Fixed-Effects-Regressionen der Berufsschulnoten (Basismodell und erweitertes Basismodell) 35
Tabelle 5: Beitrag des Multicheck-Gesamtergebnisses und der Volksschulnoten zum Bestimmtheitsmass R^2 der Fixed-Effects-Regressionen des erweiterten Basismodells 37
Tabelle 6: Fixed-Effects-Regressionen der Berufsschulnoten mit beim Mittelwert geteiltem Steigungsparameter für das Multicheck-Gesamtergebnis (Spline Regression) 38
Tabelle 7: Ordered-Probit-Regressionen der durchschnittlichen Berufsschulnoten des ersten und dritten Semesters (Basismodell und erweitertes Basismodell) 42
Tabelle 8: Mithilfe der erweiterten Ordered-Probit-Modelle nach persönlichen Merkmalen geschätzte individuelle Wahrscheinlichkeiten zum ersten, den mittleren oder dem vierten Quartil der Notenverteilung der jeweils besuchten Berufsschule zu gehören 43
Tabelle 9: Ordered-Probit-Regressionen für Multicheck-Gesamtergebnisse ab 60 Punkten (erweitertes Basismodell) 45
Tabelle 10: Schulnotenvalidität nach kantonaler Herkunft der Lernenden 49
Tabelle 11: Schulnotenvalidität nach Muttersprache der Lernenden 50
Tabelle 12: Ordered-Probit-Regressionen der durchschnittlichen Berufsschulnoten des ersten Semesters unter Verwendung der in den Schulzeugnissen ausgewiesenen entschuldigten Absenzen (zentsch01) 51
Tabelle 13: Koeffizienten einzelner über zwei Semester gemittelter Fächernoten im Ordered-Probit-Modell der Durchschnittsnoten im ersten und dritten Lehrsemester (erweitertes Basismodell) 53
Tabelle 14: Koeffizienten einzelner Multicheck-Testgebiete im Ordered-Probit-Modell der Durchschnittsnoten im ersten und dritten Lehrsemester (erweitertes Basismodell) 54
Tabelle 15: Logit- und Probit-Modell der Wahrscheinlichkeit, unentschuldigte Absenzen während den ersten drei Lehrsemestern zu verschulden 57
Tabelle 16: Probit- und Logit-Modell der Wahrscheinlichkeit einer Lehrvertragsauflösung 61
Tabelle 17: Aus den Personaldossiers erhobene Kenngrössen und deren Stata-Bezeichnung (kursiv) 76
Tabelle 18: OLS-Regressionen der Berufsschulnoten (Basismodell und erweitertes Basismodell) 80
Tabelle 19: Cluster-robuste GLS-Regressionen der Berufsschulnoten (Basismodell und erweitertes Basismodell) 80
Tabelle 20: Fixed-Effects-Regressionen des Basismodells nach Multicheckversion 81
Tabelle 21: Prädiktive Validität einzelner Semesterzeugnisse in den Fixed-Effects-Regressionen des erweiterten Basismodells 81
Tabelle 22: Ordered-Probit-Modelle der Berufsschulnoten des ersten und dritten Semesters unter Verwendung der sozialen Herkunft der Lernenden 83
Tabelle 23: Signifikanz und Effektgrösse des Multicheck-Gesamtergebnisses im Ordered-Probit-Modell unter Verwendung eines immer grösseren Satzes an Kontrollvariablen 84
Tabelle 24: Signifikanz und Effektgrösse der Schulzeugnisnoten im Ordered-Probit-Modell unter Verwendung eines immer grösseren Satzes an Kontrollvariablen 85
Tabelle 25: Fixed-Effects-Regressionen mit Interaktionstermen zwischen den durchschnittlichen Schulnoten der Bewerbungszeugnisse und dem Schultyp 86
Tabelle 26: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Deutsch 87
Tabelle 27: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Französisch 88

Tabelle 28: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Englisch .. 89
Tabelle 29: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Detailhandelskenntnisse 90
Tabelle 30: Fixed-Effects-Modelle der durchschnittlichen Berufsschulnote im Fach Branchenkunde 91
Tabelle 31: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Gesellschaft 92
Tabelle 32: Fixed-Effects-Modelle der durchschnittlichen Erst- und Drittsemester-Berufsschulnote im Fach Wirtschaft 93
Tabelle 33: Fixed-Effects-, Random-Effects-Logit- und -Probit-Modelle der unentschuldigten Absenzen..... 94

1 EINLEITUNG

1.1 PROBLEM- UND FRAGESTELLUNG

In den USA ist es seit Jahren üblich, dass externe Leistungstests Bildungsverläufe strukturieren. So hängt beispielsweise der Zugang zu höherer schulischer Bildung massgeblich von den Resultaten in einem externen Studierfähigkeitstest ab. Im Gegensatz dazu war der Einsatz standardisierter Tests in der Schweizer Bildungslandschaft lange Zeit auf die Selektionsverfahren grosser Unternehmen beschränkt. Das hat sich aber in den letzten Jahren vor allem im Übergang zwischen Sekundarstufe I und Berufsbildung geändert. Hier setzte sich die Verwendung meist privat entwickelter, computerbasierter „Berufseignungstests" immer wie mehr durch. Mittlerweile gehört die Auswertung eines „Multichecks" oder „basic-checks" zum festen Bestandteil der Bewerbungsunterlagen vieler Berufslehren. Die Tests scheinen dabei nicht nur eine ökonomischere, sondern auch eine objektivere Selektion zu erlauben, indem die Auswertungen einfache und unabhängige Vergleiche schulischer und kognitiver Fähigkeiten der Bewerber ermöglichen.

In dem Masse, wie das Gewicht von standardisierten Berufseignungstests in der (Vor-) Selektion zugenommen hat, wurden schulische Leistungsausweise darin abgelöst (Moser 2004c). Denn Berufseignungstests sind nicht zuletzt eine Reaktion der Wirtschaft auf das Fehlen eines verlässlichen „Signallings" der Kompetenzen von Schulabgängern durch die Volksschule (Bishop 1994). Der Zweifel der Wirtschaft an den schulischen Leistungsausweisen kommt nicht von ungefähr: Unterschiedliche Schulniveaus, verschiedenartig hierarchisierte Schulzüge, je nach Kanton ungleiche Lehrpläne und ständige Reformen – es ist eine Herausforderung, das schweizerische Schulsystem überhaupt zu überblicken. Ein Ausbildungsverantwortlicher, der Bewerber aufgrund ihrer effektiven schulischen und kognitiven Fähigkeiten anstellen möchte, steht vor einer sehr schwierigen Aufgabe – zumal die Schulnoten, die genau diese Informationen liefern sollten, ihren Zweck nicht erfüllen. Studien zeigen, wie zufällige und systematische Einflussfaktoren wie Lehrer, Klasse und Eltern der Jugendlichen den Signalgehalt von Schulnoten unterminieren. So kulminieren die Folgen eines „entwerteten" (Feischmann 2004, 6), für Kronig (2007, 201) ausserhalb des Klassenzimmers gar „sinnentleerten" Notensystems in einem beträchtlichen Wohlfahrtsverlust (Costrell 1994).

Tragen die Berufseignungstests dazu bei, diesen Wohlfahrtsverlust zu verkleinern? Können die Tests das mangelhafte Signalling zwischen Volksschule und Berufsbildung verbessern? Eine Beantwortung dieser Fragen ist zwangsläufig mit der Erhebung der diagnostischen und prognostischen Zuverlässigkeit der eingesetzten Tests verbunden. Doch genau diese Überprüfung fand bislang kaum statt. Die fehlende Forschungsaufmerksamkeit überrascht umso mehr, als die Berufseignungstest inzwischen eine zentrale Bedeutung in der Schweizer Lehrlingsselektion (Moser 2004a; Mühlemann et al. 2007) erhalten haben und ihr Einfluss auf Bildungs- und Lebensverläufe von jungen Menschen möglicherweise erheblich ist.

Eine ähnliche Forschungslücke zeigt sich in Bezug auf die Schulnoten. Zwar ist die Kritik an ihnen mannigfach, trotzdem belegen Studien immer wieder einen umfangreichen Einsatz der Schulnoten in

der Schweizer Lehrlingsselektion (Moser 2004c; Schmid und Storni 2004). Welche Informationen enthalten die schulischen Leistungsausweise, die sie für die Betriebe verwertbar machen? Ist der Aussagegehalt von Schulnoten tatsächlich so gering, wie es die Wirtschaft und gewisse Autoren wissenschaftlicher Studien vermuten? Oder sind die Signalwerte von Schulnoten und Schulniveau etwa besser als ihr Ruf?

Die vorliegende Arbeit trägt mithilfe einer quantitativen Fallstudie dazu bei, diese Forschungslücken zu schliessen. Anhand der ökonomischen Signalling-Theorie wird theoretisch wie empirisch untersucht, wie es um den „Signalgehalt" oder „Signalwert" des Berufseignungstests „Multichecks Detailhandel und Service"[1] und jenem von schweizerischen Volksschulnoten steht. Konkret geht die Studie den Fragen nach, inwiefern Multicheck und Volksschulnoten zur Vorhersage der späteren Berufsschulleistungen, der Wahrscheinlichkeit von unentschuldigten Absenzen in der Berufsschule, und zur Prognose der Wahrscheinlichkeit einer Lehrvertragsauflösung beitragen können. Der Hauptfokus liebt dabei aus methodischen Gründen, aber auch angesichts ihrer Relevanz und ihrer Nähe zu den interessierenden Selektionsinstrumenten auf den Berufsschulnoten.

Konzeptionell lässt sich die Studie damit als Validitätsanalyse bezeichnen (Diekmann 2005). Allerdings werden zur Beantwortung der Fragestellung nicht wie üblich univariate (z. B. Widmer 2006), sondern multivariate statistische Verfahren angewendet. Das erlaubt zu untersuchen, ob die Ergebnisse zum Signalwert der beiden Selektionsinstrumente kontextuellen Bedingtheiten unterliegen. Gewinnen die Schulnoten an Aussagekraft, wenn ihre Entstehungsumgebung berücksichtigt wird? Hängt ihr Signalgehalt vom kantonalen Schulsystem oder von der sozialen oder kulturellen Herkunft der Lernenden ab? Sind Realschulnoten ähnlich vertrauenswürdig wie jene der Sekundarschule? Und wie gross fällt der Signalwert des Multichecks aus, wenn die schulischen Leistungsausweise und andere Kontrollvariablen bereits berücksichtigt werden?

Zur Beantwortung der aufgeworfenen Fragen wurden in einer Vollerhebung die Daten aller Lernenden der Migros-Genossenschaften Aare, Luzern und Zürich erhoben, welche in den Jahren 2006 und 2007 die dreijährige Lehre als Detailhandelsfachmann beziehungsweise -fachfrau[2] aufnahmen. Insgesamt wurden so Informationen zu persönlichen Merkmalen, volksschulischen Leistungen, räumlicher und sozioökonomischer Herkunft und zu Multicheck-Ergebnissen von über 300 Lernenden erfasst, welche anschliessend in linearen Regressionsmodellen und in nichtlinearen Wahrscheinlichkeitsmodellen zur Prognose der beruflichen Erfolgskriterien – Berufsschulnoten, schulische Absenzen und Lehrvertragsauflösungen – eingesetzt wurden.

Die Resultate der Studie deuten darauf hin, dass der Detailhandels-Multicheck einen wesentlich kleineren Signalgehalt aufweist, als theoretisch zu vermuten ist, während es sich bei den Erkenntnissen zu den Schulnoten genau umgekehrt verhält. So ist der Schluss zulässig, dass Schulnoten unter Berücksichtigung kontextueller Gegebenheiten sehr wohl zur Lehrlingsselektion eingesetzt werden

[1] Wenn im Folgenden von „Multicheck" in Zusammenhang mit dem in dieser Arbeit untersuchten Berufseignungstest die Rede ist, handelt es sich um jenen Multicheck „Detailhandel und Service".

[2] In der Arbeit wird im Sinne der Leserlichkeit darauf verzichtet, konsequent eine geschlechtergerechte Sprache zu verwenden. Die weiblichen Gruppenmitglieder sind allerdings in Begriffen wie Kandidat oder Lernender explizit einbezogen.

können, während der prognostische Mehrwert des Detailhandels-Multichecks für die Migros grösstenteils – aber mit Ausnahme etwa der sprachlichen Testgebiete – durch die (schulischen) Kontrollvariablen absorbiert wird. So vermögen die Noten möglicherweise aufgrund ihrer sozialen Dimension oder aufgrund der regelmässigen Leistungsmessung, welche ihnen zugrunde liegt, den Multicheck in der Vorhersage aller drei betrachteten Kriterien des Lehrerfolgs klar zu schlagen.

Allerdings ist schwierig zu beantworten, inwiefern diese Erkenntnisse generalisierbar sind. Insbesondere die Beschränkung auf einen einzelnen Grossbetrieb dürfte die externe Validität der Untersuchung einschränken. Doch zeigen die Befunde eines deutlich: Standardisierung und Vergleichbarkeit des Eignungstests allein reichen für einen hohen Signalgehalt nicht aus. Im vorliegenden Kontext trug der Eignungstest deshalb nur wenig dazu bei, das Signalling zwischen Volksschule und Arbeitsmarkt zu verbessern.

Die Arbeit ist folgendermassen aufgebaut: Nach den einleitenden Bemerkungen zum Stand der Forschung und zum in der Arbeit untersuchten Eignungstest Multicheck, widmet sich Kapitel 2 dem Problem der Lehrlingsselektion und stellt diese in den Kontext der ökonomischen Signalling-Theorie. Das Kapitel dient dazu, Hypothesen zum Signalwert der Volksschulnoten und zu jenem des Detailhandels-Multichecks zu formulieren. Kapitel 3 stellt anschliessend die für die Beantwortung der Fragestellung erhobenen Daten vor, beschreibt das erfassungsmethodische Vorgehen und erläutert die aus den Daten gebildeten Variablen. In Abschnitt 4 werden die formulierten Hypothesen in univariaten und multivariaten statistischen Analysen überprüft. Der letzte Abschnitt liefert schliesslich einige Schlussfolgerungen zu den gewonnenen Erkenntnissen.

1.2 Stand der Forschung

Sehr gut erforscht ist der Übergang von der Schule ins Berufsleben in der Schweiz. Vor allem aus bildungssoziologischer Warte sind in letzter Zeit eine ganze Reihe von Arbeiten zur „Transition an der ersten Schwelle“ erschienen (überblickend: Häfeli und Schellenberg 2009). Die meisten Autoren widmen sich dabei dem Selektionsverhalten der Betriebe (z. B. Moser 2004c), der Sicht der Jugendlichen auf die Lehrstellensuche (Herzog et al. 2006) oder den Lehrstellenchancen der Schulabgänger (bspw. Hupka et al. 2006), wobei der Fokus meist auf Grossunternehmen gerichtet ist (Ausnahme: Imdorf 2007a).[3] Doch nicht nur die eigentliche Transition, sondern auch die Erfolgsfaktoren in der beruflichen Grundbildung sind gut erforscht (z. B. Häfeli und Schellenberg 2009). So befasste sich die Forschung mit den Gründen von Lehrvertragsauflösungen (Stalder und Schmid 2006) und den Gründen, an die Berufslehre eine Anschlusslösung zu finden (Stalder et al. 2008).

Ähnlich beschäftigt sich die bildungssoziologische und erziehungswissenschaftliche Forschung seit Jahren mit der Aussagekraft und dem Sinn der Notengebung in der Volksschule. An der Notengebung wird dabei meist kein gutes Haar gelassen (bspw. Ingenkamp 1995; Kronig 2007). Trotzdem fanden vor allem psychologische Studien immer wieder einen hohen prädiktiven Gehalt von Schulnoten in der Vorhersage von Ausbildungs- und Arbeitsmarkterfolg (z. B. Baron-Boldt 1989; Roth et al. 1996)

[3] Der Übergang von der Schule in die Berufsschule erhielt vor allem deshalb grosse Forschungsaufmerksamkeit, weil die Bildungschancen auf Sekundarstufe II stark von askriptiven Merkmalen abhängen (Häberlin et al. 2004).

und die Betriebe verwenden die schulischen Leistungsausweise in der Lehrlingsselektion (Neuenschwander und Malti 2009). Mit dieser Diskrepanz zwischen theoretisch zu erwartendem und empirisch feststellbarem Signalgehalt von Volksschulnoten befassten sich aber nur wenige ökonomische und soziologische Autoren explizit (bspw. Grant 2007; Miller 1998). Die ökonomische Forschung fokussiert dabei fast ausschliesslich auf den Zusammenhang, den sich zwischen institutioneller Begebenheit – insbesondere externer Leistungsüberprüfung – und Signalwert ergibt (De Paola und Scoppa 2010; Schwerdt und Wössmann 2009). Eine umfassende Studie, welche den Informationsgehalt von Schulnoten und dessen Bedingtheiten mit der betrieblichen Verwendung schulischer Leistungsausweise in der Selektion und dem mit guten Schulnoten verbundenen Arbeitsmarkterfolg verbindet, fehlt bislang.

Ein ähnliches Forschungsdesiderat sind Untersuchungen, welche die Aussagekraft der privaten in der Schweiz eingesetzten Berufseignungstests wissenschaftlich erforschen. Zwar gibt es umfassende Evidenz dafür, dass gut ausgestaltete Intelligenz- und Schulleistungstests einen guten prädiktiven Gehalt aufweisen (Kramer 2009; Stalder et al. 2008). Trotz den Befunden dieser Arbeiten ist bis heute weitgehend unklar, wie es um die tatsächliche Aussagekraft der Berufseignungstests wirklich steht. Die geringe Forschungsaufmerksamkeit, die beispielsweise dem Multicheck bislang zukam, überrascht angesichts der Kritik, die gegenüber solchen Verfahren immer wieder laut wurden (Galliker 2003; Jungo und Zihlmann 2002). Eine der wenigen Ausnahmen bildet hier die Lizentiatsarbeit von Widmer (2006), in welcher Reliabilität und Validität des kaufmännischen Multichecks – allerdings mit teilweise zu kleinen Stichprobengrössen – ausführlich analysiert werden. Freilich hat Multicheck – wie im Übrigen auch andere Anbieter (Jungo und Zihlmann 2002) – selber Analysen zur Validität ihres Tests durchgeführt, doch genügen diese in der verfügbaren Dokumentation wissenschaftlichen Ansprüchen nicht (Multicheck 2008; Multicheck 2010c). Trotzdem werden diese Studien in dieser Arbeit einer kritischen Überprüfung unterzogen.

1.3 Der Multicheck

Die Firma Multicheck mit Sitz in Konolfingen ist gemäss eigenen Angaben der schweizweit grösste Anbieter berufsbezogener Eignungsabklärungen und Kompetenzanalysen für Jugendliche, die ins Berufsleben einsteigen. Jährlich absolvieren mehr als 30‘000 Lehrstellensuchende den Berufseignungstest Multicheck, der zurzeit in sechs nach Berufsfeld und -branche verschiedenen Versionen existiert, darunter der kaufmännische Multicheck und der hier betrachtete Multicheck „Detailhandel und Service“. Der Test verfolgt gemäss Widmer (2006, 49) das Ziel, „lehrpraktische und insbesondere theoretische Ausbildungsleistungen von Auszubildenden vorherzusagen“.

Der Detailhandels-Multicheck ist eine reine Computerabklärung. Der Test umfasst einerseits den schulischen Bildungsstand der Berufsinteressierten. Gemäss Aussagen der Verantwortlichen der Firma Multicheck orientieren sich Fächer und Themen am Lehrplan der Berufsschule, während sich die Anforderungen nach dem Niveau der obligatorischen Schule am Ende der achten Klasse richten. Konkret erfasst der Test das Wissen in den Bereichen Deutsch, Mathematik und in einer der beiden, oder auf freiwilliger Basis in beiden Fremdsprachen (Französisch oder Englisch). Andererseits erhebt der Test auch das Potenzial der Kandidaten. Im Spezifischen enthält der Test je ein Testgebiet

in den Bereichen Logik, Merk- und Konzentrationsfähigkeit.[4] Er dauert ungefähr zwei Stunden und kostet die Kandidaten zurzeit 70 Franken. Einige Tage nach der Durchführung erhalten die Kandidaten eine Auswertung auf einer A4-Seite, welche den Bewerbungsunterlagen beizulegen ist. Diese weist einerseits ein Gesamtresultat, die Ergebnisse in den einzelnen Testgebieten und je ein Ergebnis in den Bereichen „Potenzial" und „Schulwissen" aus.[5] Schliesslich liefert die Auswertung auch Information über die für die einzelnen Gebiete benötigte Zeit und die Arbeitsqualität des Kandidaten. Diese Elemente fliessen im Übrigen auch in das Gesamtergebnis ein[6]: So wird laut Multicheck die Anzahl richtig gelöster Aufgaben in Prozent mit dem Qualitäts- und Zeitbonus zusammen in eine neue Punktzahl überführt, über welche dann eine Normalverteilung gelegt wird.[7] Die Darstellung erfolgt schlussendlich in Prozentrangwerten mit Mittelwert 50, welche dem Kandidaten Auskunft geben, wie viele der Absolventen des Tests im Schnitt besser und wie viele im Schnitt schlechter abschneiden (siehe die Musterauswertung in Anhang 7.A). Multicheck gibt als kritische Grenze zur Berufseignung den Wert 40 an. Kandidaten, die „im Gesamtresultat und in mehreren Einzelergebnissen unter dem Wert 40 [liegen], sind möglicherweise nicht in der Lage, die in der Lehre geforderten Leistungen zu erbringen" (Multicheck 2010a). Damit sei für den Kandidaten auch die Suche nach einer beruflichen Alternative angezeigt.

Der Erfolg des Multichecks und ähnlichen Berufseignungstests wie der private „basic-check" (Galliker 2003; Jungo und Zihlmann 2002) wird vor allem mit der geringen Aussagekraft der schulischen Leistungsausweise erklärt (bspw. Häfeli und Schellenberg 2009; Imdorf 2005). Zudem haben die weniger aufwändigen und damit günstigeren betriebsexternen Verfahren in den letzten Jahren je länger je mehr betriebsinterne Leistungstests abgelöst (Moser 2004c). So verlangen zahlreiche Schweizer Unternehmen verschiedener Branchen den Multicheck oder die Resultate anderer Eignungstests mittlerweise standardmässig von jedem Bewerber. Diese Tests haben neben der Vergleichbarkeit für die Betriebe den Vorteil, dass sie für die Jugendlichen eine Hürde darstellen, sich nicht wahllos in verschiedenen Berufsfeldern zu bewerben. Die Ergebnisse der Tests sollten dabei, wie Multicheck selber (Multicheck 2010d) und kritische Autoren (Berner KMU 2010; Galliker 2003; Jungo und Zihlmann 2002) betont haben, neben den Bewerbungsunterlagen, den schulischen Leistungsausweisen, dem Bewerbungsgespräch und der Schnupperlehre nur als ein Element der Bewerbung betrachtet werden. Doch dürfte den Eignungstests gerade in der Vorselektion der Bewerber in stärker formalisierten Bewerbungsverfahren, wie sie in Grossbetrieben üblich sind, eine grosse Bedeutung zukommen (Imdorf 2005; Moser 2004c).

---

[4] Eine Analyse, inwiefern die inhaltliche Konzeption des *KV*-Multichecks den Anforderungen des Berufes entspricht, findet sich in Widmer (2006).

[5] Dabei sind die Aufgaben in den Testgebieten Konzentration, Logik und Merkfähigkeit zu einem grossen Anteil für das Ergebnis im Bereich Potenzial verantwortlich, während die Ergebnisse in den Sprachfächern und der Mathematik vor allem das Resultat im Bereich Schulwissen bestimmen.

[6] Diese Bestandteile der Multicheck-Auswertung wurden zwar auch erhoben, die Auswertungen sollen aber nicht Gegenstand dieser Arbeit sein. Die Resultate können analog zu jenen in Widmer (2006) als ernüchternd bezeichnet werden.

[7] Siehe zur Normierung und Eichung des Tests die Informationen in Multicheck (2010c).

2 Die Lehrlingsselektion als Prototyp eines Signalling-Problems

2.1 Einleitung

Die hier betrachtete Transition von der Sekundarstufe I in die Sekundarstufe II – teilweise auch als „erste Schwelle" bezeichnet (Hupka et al. 2006) – hat kein einheitliches, theoretisches Analysekonzept. In der schweizerischen Bildungsforschung wird sie meist bildungssoziologisch (siehe bspw. Kronig 2007) verortet. Ökonomisch bietet sich derweil eine Betrachtung mithilfe der Signalling-Theorie an (überblickend: Spence 2002; Weiss 1995; kritisch: Bills 2003). In ihr wird das Besetzen einer Lehrlingsstelle für einen Betrieb als Investition unter Unsicherheit interpretiert. Das Wesen von Stellenbesetzungsverfahren ist es, dass der Bewerber private Informationen besitzt, beispielsweise über seine schulischen und kognitiven Fähigkeiten, welche er nicht preisgibt, die aber für den Arbeitgeber von Interesse sind (Spence 1973).

Betrachten wir ein sehr einfaches Beispiel: Angenommen es gibt nur zwei Arten von Bewerbern für eine Lehrstelle: Jene, welche die Anforderungen der Berufsschule nicht erfüllen (Gruppe A), aber sich trotzdem bewerben, und jene, welche die Anforderungen tatsächlich erfüllen (Gruppe B). Hat der Lehrmeister nun keine Möglichkeit, die Bewerber zu unterscheiden, wird er – vorausgesetzt er ist risikoavers – keinem die Lehrstelle anbieten.[8] Allerdings wird er versuchen, auf die unbeobachtbaren Berufsschulfähigkeiten der Bewerbenden mit den Instrumenten der Bewerbungsselektion rückzuschliessen (*Screening*).[9] Dazu kann der Arbeitgeber einerseits unbeeinflussbare Merkmale des Bewerbers wie beispielsweise Geschlecht und Nationalität verwenden (Spence bezeichnete diese Merkmale als *Indizes*). Der Arbeitgeber kann sich aber auch auf *Signale* abstützen. Das sind Attribute, die der Bewerber beeinflussen kann oder konnte, deren Erwerb aber mit Kosten verbunden ist – wie beispielsweise der höchste Schulabschluss oder Schulnoten eines Bewerbers. Bildung wird damit zum Selektionshilfsmittel (*Screening Device*, Arrow 1973).

Ein Signalling-Gleichgewicht tritt nun ein, wenn die Erwartungen des Arbeitgebers mit den (rationalen) Investitionsentscheidungen der Bewerber übereinstimmen: Nehmen wir an, der Arbeitgeber erwartet, dass ein Bewerber mindestens einen Notenschnitt von 5.0 aufweisen muss, damit er die Anforderungen der Berufsschule erfüllt. Wenn nun die Bewerbenden dieses Verhalten kennen oder antizipieren, stellt sich für sie die Frage, ob sie gewillt sind, den Aufwand zu betreiben, der für diese Note gefordert ist. Angenommen ein grosser Teil der Gruppe B ist bereit, die psychischen und zeitlichen Kosten zum Erreichen des Signals auf sich zu nehmen, weil es für sie netto – durch die besseren Lehrstellenchancen – rentiert. Im Gegensatz dazu sind die Kosten für sehr viele in Gruppe A aufgrund ihrer kleineren schulischen Fähigkeiten grösser, so dass sich für sie die schulische Investition

[8] Bishop (1994, 12) beschreibt eine solche Situation: „One of the saddest consequences of the lack of signals of school achievement is that employers offering training and job security are unwilling to take the risk of hiring a recent high school graduate."

[9] Hier zeigt sich die Verbindung zur *Screening*-Theorie (Stiglitz 1975): Während Stellensuchende Signale aussenden, *screenen* Arbeitgeber. Die beiden Theorien sind also komplementär.

nicht rechnet. So entsteht ein Gleichgewicht: Diejenigen, welche eine 5.0 erreicht haben, werden für ihren Aufwand belohnt und erhalten eine Lehrstelle, die anderen nicht. In einem solchen, *gepoolten* Gleichgewicht sind dann auch die ex ante adäquaten Erwartungen des Arbeitgebers erfüllt – sie werden ex post erfüllt, da ja fast alle angestellten Lernenden die Berufsschule bestehen. Die Erwartungen des Arbeitgebers sind selbst-bestätigend (*self-confirming*) (Spence 2002).

Natürlich könnte es sein, dass bereits ein Notenschnitt von 4.5 ausreichen würde, um das erwünschte gepoolte Gleichgewicht herzustellen – doch insofern der Arbeitgeber die Erwartungen gebildet hat und sie sich bestätigen, wird er seinem einmal erfolgreichen Vorgehen treu bleiben. Erwähnenswert ist zudem, dass ein solches selektionierendes Gleichgewicht nur entsteht, wenn der Erwerb des Signals mit Kosten verbunden ist, die negativ mit der versteckten Eigenschaft – den schulischen Fähigkeiten – korrelieren. Wäre dies nicht der Fall, würde auch Gruppe A das Signal erwerben und die Fraktionen wären wiederum nicht unterscheidbar (Stiglitz 1975; Spence 1973). Schliesslich können Signalling-Gleichgewichte diskriminierend sein. So entsteht unter Umständen ein für alle Seiten rationales Gleichgewicht, in welchem ein Arbeitgeber durch den Erfolg seiner männlichen Lehrlinge stets in der Erwartung bestätigt wird, weibliche Lehrlinge würden den Anforderungen der Berufsschule nicht genügen (Spence 1973).

Signalling-Gleichgewichte mögen sehr vereinfachend und stilisierend wirken und die Muster, die sie erkennen lassen, hängen von den vereinfachenden Annahmen ab, welche getroffen werden. Trotzdem sind die Erwartungsbildung über die Fähigkeiten der Bewerber auf Seiten des Arbeitgebers, wie auch das Erwerben von Bildungsabschlüssen als Signale der eigenen Produktivität auf Seite der Bewerber, unbestreitbar Kernelemente jedes Stellenbewerbungsverfahrens. Und wenn Imdorf (2007c, 100) beschreibt, wie „ausländisch geltende Jugendliche fälschlicherweise unter Verdacht [stehen], betriebliche Extraprobleme zu verursachen“ und „erst die Bereitschaft, mit ‚ausländischen‘ Jugendlichen überhaupt in Kontakt zu treten, es […] den Betrieben [ermöglichen würde], solche Annahmen im Praxistest zu widerlegen“, dann kommt dies dem beschriebenen, diskriminierenden Gleichgewicht sehr nahe.

In unserem Zusammenhang interessiert nun die Frage, wie verlässlich die Signale „Schulnoten“ und „Multicheck“ sind. In der Literatur wird von *Signaling Value* (bspw. Schwerdt und Wössmann 2009) geredet. Im Folgenden werden deshalb die Begriffe „Signalwert“ und „Signalgehalt“ verwendet. Aus der obigen Beschreibung des Gleichgewichtes wird ersichtlich, dass sowohl Studien von Interesse sind, welche den statistischen Zusammenhang zwischen den interessierenden Variablen und dem „Lehrerfolg“ ermitteln, wie auch Studien, die den Einsatz von Schulnoten und Multicheck in der Selektionsentscheidung aufzeigen – schliesslich beruht dieses Verhalten ja auf den sich bestätigenden Erwartungen der Selektionsverantwortlichen.

2.2 Theoretische Überlegungen zum Signalgehalt schulischer Leistungsnachweise

Im Folgenden wird für zwei schulische Formalqualifikationen – Schulnoten (Kapitel 2.2.1) und das schulische Anforderungsniveau (Kapitel 2.2.2) – der theoretisch zu erwartenden Signalgehalt ermittelt. Die in den meisten schweizerischen Volksschulzeugnissen enthaltenen Informationen zum

Betragen und zu den Absenzen der Schüler sind weitere, eventuell im Arbeitsmarkt verwertbare Signale, die den Signalgehalt des Schul*zeugnisses* beeinflussen. Der Signalgehalt dieser Elemente der Volksschulzeugnisse wird in Abschnitt 4.4.3 diskutiert.

2.2.1 Volksschulnoten als Signal

Ökonomische Studien analysieren den Signalwert von Schulnoten innerhalb von Messfehlermodellen. Diesen Weg beschreiten beispielsweise De Paola und Scoppa (2010), Grant (2007) und Schwerdt und Wössmann (2009). Die Modelle dieser Autoren verstehen die Schulnote beziehungsweise die Schulqualifikationen als eine Linearkombination von produktiven Fähigkeiten und einem oder mehreren Fehlertermen verschiedenen Ursprungs. Diese Modelle vermögen aber im vorliegenden Kontext nur wenig Substanzielles beizutragen, da sie stark auf die institutionellen Bedingtheiten des Signalwerts von Schulnoten – insbesondere die externe Leistungsüberprüfung – fokussieren. Dem umfassenden Charakter des untersuchten Signalling-Problems werden sie deshalb nicht gerecht.

Interessanter ist es deshalb, die Wichtigkeit von Schulnoten in der Lehrlingsselektion zu diskutieren, wie sie empirische Studien etablieren. Allerdings liefert die Forschung auf die Frage, welche Bedeutung Schulnoten in der Lehrlingsselektion in der Schweiz zukommt, keine klare Antwort. So hängen die Lehrstellenchancen eines Bewerbers in einem Grossbetrieb gemäss der grossangelegten Studie von Moser (2004c) nur sehr eingeschränkt von seinen Volksschulnoten ab. Die Auswertungen von Hupka et al. (2006) mithilfe der Daten des Jugendlängsschnitts TREE deuten jedoch darauf hin, dass ein guter Notendurchschnitt in den Zeugnissen der Volksschule einen grossen positiven Einfluss auf die *Eintrittsrate* in eine zertifizierende Sek-II-Ausbildung haben. Dieses gemischte Bild bestätigt sich auch in den untersuchten Migros-Genossenschaften: Je nach Lehrlingsverantwortlichem werden Schulnoten hoch oder nur gering gewichtet.

Die Schulnoten werden dabei gemäss Aussagen der Lehrlingsbetreuer vor allem zur Vorselektion eingesetzt. Später treten sie in ihrer Bedeutung für die Selektionsentscheidung deutlich hinter die Eindrücke aus der Schnupperlehre und dem Vorstellungsgespräch zurück. Das entspricht grob dem Bild, das Studien und Umfragen zur Bedeutung schulischer Leistungsausweise in der Lehrlingsselektion der Schweiz zeichnen (Häberlin et al. 2004; Neuenschwander und Wismer 2010; Schmid und Storni 2004; Stalder 2000; Stalder und Stricker 2009). Schulnoten erfüllen dabei – auch für die Lehrlingsverantwortlichen der Migros – vor allem zwei Funktionen: Einerseits dienen sie angesichts der vielen Bewerbungen dazu, das Selektionsverfahren zu ökonomisieren. Andererseits werden Schulnoten vor allem zur Einschätzung der Berufsschultauglichkeit und damit zur Verhinderung berufsschulbedingter Problemen in der Lehre eingesetzt (Imdorf 2007a; Imdorf 2009).[10]

[10] Asthon und Maguire (1980) unterscheiden insgesamt fünf Funktionen von Schulnoten in der Lehrlingsselektion, von denen Imdorf (2009) in Schweizer KMUs vier von fünf beobachtet: determinierende Funktion (minimales Schulnotenniveau wird gefordert, jedoch wird der Kandidat mit den besten Schulqualifikationen eingestellt); Screening-Funktion (Schulnote dient der Vorselektion); fokussierende Funktion (vorteilhafte Schulqualifikationen sind erwünscht, aber wenn ein Kandidat durch nicht-akademische Qualitäten überzeugt, verliert ein unvorteilhaftes Schulzeugnis an Bedeutung); negative Funktion (sehr gute Schulqualifikationen sind ein Ausschlusskriterium); Funktionslosigkeit. Einzig für die determinierende Funktion findet Imdorf keine Belege.

Insgesamt zeigt sich aber ein ambivalentes Bild der betrieblichen Verwertung von Schulnoten: Zwar werden sie in der Selektion – vielleicht aufgrund ihrer einfachen Verfügbarkeit – sehr oft eingesetzt. Andererseits kommt ihnen in der Lehrlingsselektion schweizerischer Betriebe nur sekundäre und kontextabhängige Bedeutung zu (Imdorf 2007a). So äussern die Verantwortlichen in den Unternehmen – im Übrigen auch jene der Migros – auch immer wieder Zweifel an der Aussagekraft und Vergleichbarkeit der Noten (Imdorf 2009; Moser 2004c). Diese Zweifel sind aus wissenschaftlicher Sicht gerechtfertigt: Die Korrelation zwischen Schulnoten und objektiven Schulleistungstests ist erwiesenermassen sehr gering (Baron-Boldt 1989; Kronig 2007; Leiter und Brown 1985; Moser 2004c; Ausnahme: Neuenschwander und Malti 2009).

Gründe für dieses Auseinanderklaffen von effektiver Leistungsfähigkeit und dem durch die Note signalisierten Leistungsstand gibt es viele.[11] So haben Schulstufe, Schulart, die Grösse des Schulortes und die räumliche Herkunft nachweislich Einfluss auf die Notengebung (Baron-Boldt 1989). Der wichtigste Verzerrungsfaktor für den Signalgehalt der Noten ist aber zweifelsohne der Lehrer, dessen Leistungsdiagnose von seinen persönlichen Merkmalen (z. B. Alter, Erfahrung, Geschlecht, Schulfach und Urteilstendenzen) und weiteren Zufälligkeiten (oft irrelevante Fähigkeiten des Schülers wie z. B. Schrift und Grammatik) abhängt. Die grössten lehrerbedingten Verzerrungen erhält die Note jedoch aufgrund von Referenzgruppenfehlern. Diese entstehen, weil sich Lehrer bei der Notengebung nicht an objektiven Kriterien, sondern vielmehr an einer sozialen Bezugsnorm – der Klasse und der Schule – orientieren. So werden Schüler in starken Schulklassen zuweilen stark benachteiligt (Baron-Boldt 1989; Kronig 2007; Leiter und Brown 1985; Resh 2009). Daneben haben die persönlichen Merkmale der Schüler einen erheblichen Einfluss auf die Schulnoten. So ist mehrfach gezeigt worden, dass Schülerinnen bessere Noten als Schüler erhalten und dass ein tiefer sozioökonomischer Status mit schlechteren Schulnoten einhergeht (Häberlin et al. 2004; Kronig 2007; Lekholm und Cliffordson 2009; Duckworth und Seligman 2006; Perrig-Chiello et al. 1997).[12]

Trotz all dieser Zufälligkeiten und systematischen Verzerrungen (Kronig 2007) attestieren Studien, die den Zusammenhang zwischen Schulnoten und diverser Erfolgskriterien untersuchen, den Schulnoten einen erstaunlich grossen Prognosegehalt. So weisen Studien in verschiedenen Ländern beachtliche Korrelationen zwischen Volksschulnoten und dem Ausbildungserfolg nach (Baron-Boldt 1989; Roth et al. 1996; Schuler et al. 1990; Stamm et al. 2009).[13] Zudem scheinen bessere Schulnoten zumindest in den USA langfristig mit einem höheren Lohn einherzugehen (Miller 1998; Schwerdt und Wössmann 2009).[14]

---

[11] Vgl. Ingenkamp (1995) oder Kronig (2007) für eine umfassende Darstellung.

[12] Ein weiterer Effekt, der die Aussagekraft von Schulnoten unterminiert und dem gerade in der ökonomischen Literatur viel Aufmerksamkeit zuteil kam, ist das Problem der Noteninflation (vgl. Chan et al. 2007; Johnson 2003; Millman et al. 1983).

[13] Auch in Bezug auf den Studienerfolg sind Schulnoten prädiktiv valide: So sind laut den Meta-Analysen von Baron-Boldt (1989) und Trapmann et al. (2007) deutsche Abiturnoten der beste Einzelprädiktor des Studienerfolgs. Ähnliche Resultate finden sich für andere Länder: z. B. Cohn et al. (2004) für die USA oder Schuler et al. (1990) für Frankreich.

[14] In der kurzen Frist ist die Evidenz für die USA allerdings uneinheitlich. So haben umfassende Forschungsprogramme Zweifel an der betrieblichen Verwertung von Schulnoten (Miller und Rosenbaum 1997) und am

Wie lässt sich die Divergenz zwischen Vergleichbarkeit der Noten und ihrem Vorhersagegehalt erklären? Einerseits verwenden die zitierten Studien oftmals Mittelwerte zur Prognose gebildet. Dies verringert – in den Messfehlermodellen gesprochen – insbesondere das Gewicht idiosynkratrischer Fehler, die beispielsweise durch eine schlechte Tagesform des Schülers zustande kommen, und den Einfluss jener Fehler, die auf den Lehrer zurückzuführen sind (Baron-Boldt 1989; De Paola und Scoppa 2010; Grant 2007).

Die zweite Erklärung der Divergenz wurde bereits von Baron-Boldt (1989, 21) angedeutet: „Weiterhin sind Schulnoten zwar ein Mass der allgemeinen intellektuellen Leistungsfähigkeit, aber sie beinhalten gleichzeitig auch noch andere Aspekte der Persönlichkeit. […] Die Kriterien bei der Prognose von Ausbildungs- und Studienerfolg sind häufig ebenfalls Noten […]. Damit wäre es denkbar, dass die Verbindung aus Leistungs- und Persönlichkeitsbeurteilung, […] im Hinblick auf den Zweck der Ausbildungs- und Studienerfolgsprognose gar nicht so negativ zu werten ist." Mit anderen Worten: Schulnoten messen zwar die tatsächliche Leistung oder die allgemeine und fachspezifische Kapazität (Intelligenz- und Begabungsfaktor, siehe z. B. Laidra et al. 2007) nur ungenau, doch enthalten sie eine andere Komponente, welche ihrem prädiktiven Gehalt zuträglich ist. Das zweite Element kann als Anpassungsfaktor charakterisiert werden, der von Persönlichkeit (vor allem Offenheit, Freundlichkeit und Pflichtbewusstsein), Interesse und Motivation abhängt (Baron-Boldt 1989; Imdorf 2007a; Lekholm und Cliffordson 2009; Miller 1998; Poropat 2009). Den Aussagen der Migros-Verantwortlichen und der in der Studie von Imdorf (2007a) befragten Arbeitgeber nach zu urteilen, wird die soziale Dimension der Notengebung bei der Selektionsentscheidung berücksichtigt.

Diese letzte Erkenntnis hat Konsequenzen für die ökonomische Betrachtung des Signalwerts von Schulnoten. Sie führt dazu, dass eine intensive theoretische Betrachtung des Schulnotensignalgehalts innerhalb konventioneller Signalling-Modelle nur bedingt erfolgsversprechend wäre, geht die Signalling-Theorie doch gemeinhin von einer eindimensionalen Fähigkeit aus, welche durch das Signal vermittelt wird. Wie Heckman und seine Koautoren ausführten (Heckman und Vytlacil 2001; Heckman et al. 2006), erfordert die Modellierung gemischter Signale eine umfassende Umformulierung der Signalling-Theorie. So ist der theoretisch weniger umfassende Weg, den Signalgehalt von Schulnoten innerhalb von Messfehlermodellen zu analysieren, vorläufig erfolgsversprechender.

2.2.2 Der Schultypus als Signal

Neben den Schulnoten ist der auf Sekundarstufe I absolvierte Schultyp das wichtigste Signal schulischer Leistungsfähigkeit.[15] Der Signalgehalt, der dem besuchten Schultypus von den schweizerischen Ausbildungsbetrieben attestiert wird, ist nicht zu unterschätzen: „Der Schultyp ist [..] bedeutsam, weil er von Schulen und Lehrbetrieben als Signal für ein bestimmtes Leistungs- bzw. Qualifikations-

---

Zusammenhang zwischen Schulnoten und unmittelbarem Lohn nach dem Berufseinstieg genährt (Grant 2007; Weiss 1995).

[15] Der Schultyp wird in dieser Arbeit als Signal und nicht als Index interpretiert, da er von den Jugendlichen beeinflusst werden konnte. Dies ist eine optimistische Betrachtung, wenn man berücksichtigt, dass die Transition von Primarschule in die Sekundarstufe I sehr stark auf askriptive Merkmale der Schüler zurückgeführt werden kann (z. B. Häfeli und Schellenberg 2009). Zudem ist fragwürdig, dass die Selektion in die Sekundarstufe I von den Sechstklässlern als bewusste „Investition" in das an den Arbeitsmarkt vermittelte Kompetenzniveau wahrgenommen wird, wie es die Signalling-Theorie den Bildungsentscheidungen unterstellt.

niveau interpretiert wird (Hupka et al. 2006, 10)." Dementsprechend belegen Studien zum schweizerischen Lehrlingsmarkt einen mittleren bis grossen Zusammenhang zwischen dem besuchten Schulniveau und der Chance, eine Lehrstelle zu erhalten (Häberlin et al. 2004; Hupka et al. 2006; Imdorf 2007b; Meyer 2009; Moser 2004c).

Auch die Zusammenhänge, welche Schulniveau und Kenngrössen des Lehrerfolgs aufweisen, sprechen für den Signalgehalt des auf der Sekundarstufe I besuchten Schultyps: So beeinflusst er die Chancen für einen erfolgreichen Wiedereinstieg nach einer Lehrvertragsauflösung (Schmid und Stalder 2008), die Wahrscheinlichkeit, sechs Jahren nach Schulabschluss keinen nachobligatorischen (Sek-II-) Abschluss erworben zu haben, oder die Wahrscheinlichkeit, sechs Jahre nach Schulabschluss eine Berufsmaturität absolviert zu haben (Bertschy et al. 2008).

Es macht also durchaus Sinn, dass die Betriebe den Schultypus als Signal der Leistungsfähigkeit eines Bewerbers berücksichtigen. Allerdings finden sich auch Effekte, die den Signalwert des Schultyps einschränken: So führen die häufigen Schulreformen und Umstrukturierungen im Schulwesen dazu, dass die Betriebe Mühe haben, die jeweiligen Schulniveaus voneinander zu unterscheiden (Imdorf 2009; Schmid und Storni 2004).

Interessant ist der Schultyp als Selektionskriterium in unserem Kontext auch deshalb, weil Berufseignungstests dazu beitragen könnten, die systematische Benachteiligung von Schülern aus Schulen mit Grundansprüchen (Realschülern) in der Lehrlingsauswahl zu verkleinern (Häfeli und Schellenberg 2009). Denn die Zuteilung in die Anforderungsniveaus der Sekundarstufe I ist massgeblich durch den Klassenkontext und durch zuschreibbare Merkmale wie sozioökonomischer Hintergrund der Eltern und Geschlecht erklärbar – sie hängt nur in kleinem Ausmass von den effektiven schulischen Fähigkeiten der Schüler ab (Häberlin et al. 2004; Häfeli und Schellenberg 2009; Hanushek und Wössmann 2006; Neuenschwander und Malti 2009). Dies führt dazu, dass sich die effektiven Schulleistungen der Schüler verschiedener Schulniveaus stark überlappen (BFS 2005b; BFS 2009b; SKBF 2010). Da aber der Schultyp in der Lehrlingsselektion eine teilweise stigmatisierende Wirkung hat – für gewisse Berufslehren und schulische Wege werden Realschüler gar nicht erst in Betracht gezogen oder sie erfüllen die Anforderungen nicht (Moser 2004b; Mühlemann et al. 2007) – führt die unfaire Selektion in die und die geringe Leistungstrennschärfe zwischen den Schultypen zu einer systematischen Benachteiligung der Jugendlichen in Realschulen. Insbesondere liegt der Schluss nahe, dass der sozioökonomische Hintergrund eines Jugendlichen via Schultyp einen indirekten Einfluss auf die Transition an der ersten Schwelle ausübt (BFS 2005b; Imdorf 2005; Moser 2004c; Hupka et al. 2006; Kronig 2007).

Das Problem entsteht vor allem auch deshalb, weil die Volkschule keine adäquaten Signale effektiver Leistungsfähigkeit liefert. Die Betriebe haben keine Möglichkeit, ihre Selektionsentscheidung auf die effektiven Fähigkeiten eines Bewerbers abzustellen. Sie versuchen deshalb durch die Verwendung des Schulniveaus als Selektionsinstrument das Risiko zu minimieren, einen Lernenden ohne Mindestkompetenzen anzustellen. Denn obwohl die Leistungsunterschiede zwischen den Schulniveaus in den PISA-Ergebnissen relativ gering sind, ist der Anteil der Jugendlichen, welche in Schulleistungstests nicht über elementare mathematische Kompetenzen und Lesekompetenzen verfügen, in

den Schulen mit Grundanforderungen (Realschulen) doch einiges höher, als in den Schulen mit erweiterten Anforderungen (Sekundarschulen).

2.3 Der Multicheck als Signal

Das dargestellte Informationsdefizit der Betriebe über die effektiven Fähigkeiten der Bewerber ist derart frappant, dass die effektiven Fähigkeiten eines Jugendlichen seine Lehrstellenchancen kaum beeinflussen (Häberlin et al. 2004; Hupka et al. 2006).[16] Hier springen die standardisierten Berufseignungstests in die Bresche. Sie stellen die privatwirtschaftliche Lösung des Signalling-Problems dar, indem sie den Betrieben Auskunft über die effektiven fachlichen Kompetenzen der Bewerber geben. So schreibt Moser (2004a, 46): „Weil die Schule keine einheitliche und für die Arbeitswelt verbindliche Leistungsmessung vollzieht und die Schulzeugnisse zu wenig Auskunft über den Leistungsstand geben, verlassen sich die untersuchten Grossbetriebe bei der Lehrlingsauswahl in einer ersten Runde des Selektionsverfahrens auf eigene Leistungstests, deren Ergebnisse den Ausschlag geben, ob die Jugendlichen zu einem Assessment oder Bewerbungsgespräch eingeladen werden." Der grosse Trumpf der Eignungstests ist dabei ihre Standardisierung. Die Tests erlauben eine von Lehrer, Schule, Schultypus und Kanton unabhängige Leistungsbeurteilung. Sie ergänzen oder lösen die Schulnoten in der Vorselektion ab. Zudem ermöglichen sie eine Ökonomisierung der Lehrlingsauswahl (Moser 2004c).

Kein Wunder also haben sich Berufseignungstests wie der Multicheck in der Lehrlingsselektion auf breiter Front etablieren können.[17] Die grossen Arbeitgeber des Detailhandels in der Schweiz verlangen den Multicheck mittlerweile unisono. Schon allein die grosse Bedeutung des Multichecks in der Detailhandelsbranche spricht für den Signalgehalt des Detailhandel-Multichecks. Zudem liefern Studien aus verschiedensten Bereichen empirische Belege, die für den Einsatz der Berufseignungstests und deren Aussagekraft sprechen. So existiert mittlerweise eine beträchtliche ökonomische Literatur, die darlegt, wie eine externe Leistungsüberprüfung nicht nur den Signalwert und damit die Belohnung guter schulischer Leistungen, sondern generell die Anreize der Akteure im Bildungswesen verbessert (Bishop 1997; Bishop und Wößmann 2004; Schwerdt und Wössmann 2009; Wössmann 2005) und die soziale Wohlfahrt erhöht (Costrell 1994).

Der Signalgehalt des Multichecks scheint auch deshalb gross, weil er kognitive Fähigkeiten erfasst und Studien immer wieder einen grossen Zusammenhang zwischen den von Tests gemessenen kognitiven Fähigkeiten und dem späteren Lohn belegt haben. Im Arbeitsmarkt sind die von den Tests gemessenen Fähigkeiten also nicht nur gefordert, sondern auch verwertbar (Constant und Konstantopoulos 2003; Grant 2007; Heckman und Vytlacil 2001; Kramer 2009; Murnane et al. 2001; Murnane et al. 2000). Ähnlich grosse Zusammenhänge zwischen (kognitiven) Leistungstests und

[16] Hupka et al. (2006, 22 f.) schätzen, dass sich „die faktisch bestehenden Leistungsunterschiede bezüglich Lesekompetenzen, mathematischen und wissenschaftlichen Kenntnissen […] lediglich zu etwa je 30 Prozent auch in entsprechenden Unterschieden in den schulischen Leistungsbewertungen […] niederschlagen".

[17] Die Befunde sind allerdings nicht einheitlich. Einige Studien konstatieren eine relativ geringe Bedeutung der Eignungstests in der Selektion (Neuenschwander und Wismer 2010; Schmid und Storni 2004; Stalder 2000). Vor allem grosse Betriebe (Moser 2004c; Mühlemann et al. 2007) und einzelne Branchen wie der in dieser Arbeit betrachtete Detailhandel setzen allerdings stark auf (meist externe) Leistungstests (Neuenschwander und Wismer 2010).

Arbeitsmarkterfolg liefern Studien zum schweizerischen Lehrlingsarbeitsmarkt (Häfeli und Schellenberg 2009). Untersuchungen haben zudem gezeigt, dass auch PISA-Leistungstests, die schulische Kompetenzen messen, den Erfolg zu späteren Laufbahnzeitpunkten vorhersagen (Bertschy et al. 2008; 2004; Stalder et al. 2008). Unter der Voraussetzung, dass der Multicheck, ähnlich wie die in Moser (2004b) betrachteten Eignungstests, stark mit den PISA-Leistungstests korreliert, sprechen auch diese Resultate für den Signalgehalt des Multichecks.

Doch es gibt auch Befunde, die den Signalwert des Tests möglicherweise relativieren. So legen die Ergebnisse von Stamm et al. (2009) nahe, dass überdurchschnittliche begabte Berufslernende nur zu Beginn der Lehre bessere praktische Leistungen erbringen und dass für die Leistungen in der Berufsschule nicht die Begabung entscheidend ist. Gewisse Autoren äusserten ausserdem Bedenken an der testtheoretischen Güte der in der Schweiz eingesetzten Eignungstests. So schreibt Moser (2004c, 25): „Geradezu gedankenlos ist der Umgang der Anbieter von zentral organisierten Eignungstests mit der Objektivität und der Validität der Tests. […] Die Tests können [..] zu irgendeinem Zeitpunkt im Jahr gemacht werden. Aus den internationalen Schulleistungsvergleichen ist jedoch bekannt, dass die Schulleistungen ganz entscheidend vom Alter und von der Anzahl besuchter Schuljahre abhängen. Zudem sind insbesondere in den mathematischen Fachbereichen auf der Sekundarstufe I zwischen den Lehrplänen und den Lehrmitteln kantonale Unterschiede auszumachen, die sich auf die Ergebnisse in den Eignungstests auswirken können.“[18]

Insgesamt dürfte dem Multicheck ein beachtlicher Signalgehalt zukommen, sollte der Test nicht unter einer fehlenden testtheoretischen Güte leiden. Doch eine erste wissenschaftliche Untersuchung des *kaufmännischen* Multichecks im Rahmen einer Lizentiatsarbeit an der Universität Bern fördert gemischte Ergebnisse zutage (Widmer 2006). So ist der Zusammenhang zwischen Multicheck-Ergebnis und Berufsschulnoten gerade für den Computerteil des Tests[19] relativ gering und der Test vermag die praktischen Lehrleistungen – allerdings bei einer kleinen Stichprobe – nicht vorherzusagen. Zudem weist der Multicheck eine geringe Retestreliabilität auf.[20] Insgesamt folgert Widmer aber, dass gerade der Gesamttestwert ein guter Prädiktor des Berufsschulnotendurchschnitts ist. Ähnliche Schlussfolgerungen macht auch eine Studie von Multicheck für den Detailhandels-Multicheck. Die Autoren ermitteln einen Korrelationskoeffizienten zwischen Berufsschulnotenmittelwert und Multicheck-Gesamtergebnis von 0.4 – ein achtbares Resultat in Anbetracht der zwischen Multicheck-Testdurchführung und Lehrabschlussprüfung verflossenen Zeit (Multicheck 2008; Multicheck 2010c).

[18] Moser (2004b, 22) findet denn auch einzelne „grobe Fehlurteile“ der Eignungstests, wenn er die in den PISA-Kurztests ermittelten schulischen Kompetenzen der untersuchten Jugendlichen als Vergleichsgrösse heranzieht.

[19] Im Gegensatz zum Detailhandels-Multicheck besteht der kaufmännische Multicheck aus zwei Teilen: einem Papier- und Bleistiftteil und einem Computerteil.

[20] Das heisst, dass die Ergebnisse, die beim Wiederholen des Tests erzielt wurden, mit den Ergebnissen aus dem ersten Versuch nur geringfügig übereinstimmen.

2.4 INDIZES – UNBEEINFLUSSBARE DETERMINANTEN DES LEHRERFOLGS

Viele Studien belegen, dass sich Betriebe in der Lehrlingsselektion auch auf Indizes, also für die Jugendlichen unbeeinflussbare persönliche Merkmale, abstützen. Zum Teil lässt sich dieses Verhalten dadurch erklären, dass gewisse Indizes mit unbeobachtbaren Eigenschaften der Lernenden korrelieren, welche wiederum den Lehrerfolg beeinflussen. Teilweise offenbart sich dabei aber auch diskriminierendes Verhalten der Betriebe. Da diese Einflussfaktoren die Resultate der multivariaten Analyse dieser Studie beeinflussen könnten (siehe bspw. Cohn et al. 2004; Constant und Konstantopoulos 2003; Hupka et al. 2006; Smith und Naylor 2005), werden die im schweizerischen Lehrlingsmarkt wichtigsten Indizes hier kurz besprochen. Es sind dies das Geschlecht, der familiäre, kulturelle und der sozioökonomische Hintergrund und die räumliche Herkunft. Andere Einflussfaktoren auf den Erfolg in der Lehre und damit potentielle Indizes oder Signale für die betriebliche Selektion gibt es viele. Häfeli und Schellenberg (2009) finden insgesamt 47 Einflussfaktoren auf den Erfolg in der beruflichen Ausbildung.[21]

Das Geschlecht übt neben dem indirekten Einfluss auf die Lehrlingsselektion, der sich via bessere Schulnoten und damit höheres Schulniveau manifestiert, auch einen direkten Effekt aus: Frauen haben grössere Mühe, eine Lehrstelle zu finden. Sie können also ihren schulischen Vorsprung nicht in bessere Lehrverhältnisse, Löhne oder spätere Berufsverhältnisse ummünzen (Bertschy et al. 2008; Häberlin et al. 2004; Häfeli und Schellenberg 2009; Hupka et al. 2006; Moser 2004c). Dieses Verhalten der Betriebe ist insofern überraschend, als Frauen eine Lehre weniger oft abbrechen (Bessey und Backes-Gellner 2008).

Die Bedeutung der sozialen und kulturellen Herkunft für den Erfolg im Schweizer Bildungssystem ist in den PISA-Berichten hinlänglich dokumentiert (BFS 2005b; BFS 2009a). Dabei werden zur sozialen Herkunft einerseits die finanziellen Mittel der Eltern und deren Bildung, andererseits auch die materielle Ausstattung an kulturellen und sozialen Güter (Beziehungen) gezählt. Die kulturelle Herkunft wird derweil meist durch den Migrationsstatus und die Erstsprache – hier als Muttersprache bezeichnet – bestimmt (BFS 2005b). Ähnlich wie beim Geschlecht manifestiert sich auch beim Einfluss der sozioökonomischen Herkunft ein indirekter Effekt auf den Erfolg im Lehrlingsmarkt. So führen das erzieherische Verhalten und die ungünstigen Bildungserwartungen der Eltern[22], die unvorteilhafte Zusammensetzung der Schulklasse, in der sich die Kinder befinden[23], und die Diskriminierung in der Notengebung und der Selektionsentscheidung nach der Primarschule dazu, dass Kinder und Jugendliche aus bildungsfernen Familien bereits schlechte Voraussetzungen für die Lehrstellensuche mitbringen (Häfeli und Schellenberg 2009; Kronig 2007; Moser 2004c).

---

[21] Darunter figurieren insbesondere auch individuelle Merkmale wie Persönlichkeitseigenschaften und Sozialkompetenzen, eine gute Beziehung zum Ausbildenden aber auch Einflussfaktoren aus den Bereichen Freizeitgestaltung, Demographie, wirtschaftliche Lage oder Gesellschaft.

[22] Eindrücklich sind hierzu die Erkenntnisse von Neuenschwander und Malti (2009). Sie zeigen, dass der Einfluss des sozioökonomischen Hintergrunds der Eltern auf die Bildungskarrieren der Kinder wohl vor allem über die elterlichen Bildungserwartungen vermittelt wird. Insgesamt treffen in ihrer Untersuchung 78.2 Prozent der Elternerwartungen bezüglich der Selektion in die Sekundarstufe I zu.

[23] Nicht zuletzt spielen hier also *Peer*-Effekte eine grosse Rolle (Hanushek et al. 2003; Kronig 2007; Sund 2009; Zimmerman 2003).

Zu den bestehenden Ungleichheiten kommen anschliessend direkte Effekte auf die Bildungschancen auf Sekundarstufe II, die einerseits dadurch vermittelt werden, dass Kinder aus sozial schlechteren Schichten vermehrt Bildungswege einschlagen, die unter ihren Möglichkeiten liegen (Häberlin et al. 2004; Herzog et al. 2004), die sich aber vor allem in nach sozioökonomischer Herkunft unterschiedlichen Lehrstellenchancen niederschlagen (Bertschy et al. 2008; Häberlin et al. 2004; Häfeli und Schellenberg 2009; Hupka et al. 2006; Imdorf 2005; Imdorf 2007b; Moser 2004c; Stalder et al. 2008). Das teilweise diskriminierende Verhalten der Betriebe führt denn auch zu ungleichen Bildungschancen auf Sekundarstufe II (vgl. Häfeli und Schellenberg 2009). So erhöht ein hoher Berufsstand der Eltern die Chance mit einer Berufsmaturität abzuschliessen klar, während die Gefahr, keinen nachobligatorischen Sek-II-Abschluss zu erhalten, für einen Immigranten der ersten Generation deutlich grösser ist als für einen Schweizer (Stalder et al. 2008). Die Evidenz, ob ein tiefer sozioökonomischer Status die die Wahrscheinlichkeit einer Lehrvertragsauflösung erhöht, ist allerdings gemischt (Bessey und Backes-Gellner 2007; Moser et al. 2008).

Schliesslich beeinflusst auch die räumliche Herkunft den Erfolg in der nachobligatorischen Bildung. So zeigen die Ergebnisse von Amos et al. (2009), dass die Erfolgschancen an der Lehrabschlussprüfung signifikant nach Kanton, in dem sich die Berufsschule befindet, variieren. Dies mag einerseits auf unterschiedliche Benotungspraktiken der Berufsschulen zurückzuführen sein. Andererseits machen sich hier möglicherweise die beträchtlichen kantonalen Differenzen im Bereich der obligatorischen Schule und die damit verbundene, unterschiedliche Vorbildung der Lernenden bemerkbar. Die räumliche Herkunft übt aber auch durch die sozioökonomische Zusammensetzung der Herkunftsgemeinde und der besuchten Schule indirekt einen Effekt auf die Erfolgschancen in der Sekundarstufe II aus (Betts und Morrell 1999; BFS 2005b; SKBF 2010; Smith und Naylor 2005).

2.5 „Lehrerfolg“ als Zielgrösse

Ziel dieser Studie ist es, den Signalwert von Volksschulnoten mit jenen des Detailhandels-Multichecks zu vergleichen. Hierzu sind Erfolgskenngrössen nötig, mit welchen der Signalgehalt der beiden Selektionsinstrumente ex post bestimmt werden kann. Die Zielgrössen (*Outcome*-Variablen), welche im Rahmen dieser Arbeit untersucht werden, sind die theoretischen Lehrleistungen – das heisst die Berufsschulnoten –, die Wahrscheinlichkeit einer Lehrvertragsauflösung und die Wahrscheinlichkeit unentschuldigter Absenzen in der Berufsschule derjenigen Bewerber, welche eine Lehrstelle im Detailhandel erhalten haben. Die Kriterien weisen einen inneren Zusammenhang auf: So sind schulische Leistungsprobleme der häufigste Grund für eine Lehrvertragsauflösung (Stalder und Schmid 2006). Unentschuldigte schulische Absenzen auf der anderen Seite sind oftmals Ausdruck mangelnder Motivation, schulischer Probleme oder von fehlendem Pflichtbewusstsein (Lounsbury et al. 2004) und sind häufig mit schlechteren schulischen Leistungen verbunden. Da verwundert es wenig, dass auch sie am Ursprung einer Lehrvertragsauflösung stehen können (Moser et al. 2008). Nicht zuletzt hängen die Berufsschulnoten oft auch eng mit den in dieser Studie nicht

berücksichtigten, *betrieblichen* Leistungen der Lernenden zusammen (Stamm et al. 2009; Imdorf 2009).[24]

Wie lässt sich die Verwendung dieser drei Kriterien theoretisch legitimieren? Einerseits werden sie von den Lehrlingsausbildnern als relevant erachtet und ex post registriert, woraus Erwartungen für die nächsten Bewerbungsverfahren gebildet werden. Sie sind somit im Sinne des oben beschriebenen Signalling-Gleichgewichts relevant für den Signalgehalt, welche dem Multicheck und den Schulnoten vom Selektionsverantwortlichen in der folgenden Auswahlrunde beigemessen werden (Spence 1973). Auf der anderen Seite korrelieren die betrachteten Zielgrössen teils eng mit den von Häfeli und Schellenberg (2009, 15) als Kenngrössen des „Erfolgs in der Berufsausbildung" bezeichneten Indikatoren „Durchhalten in der Lehre", „erfolgreicher Lehrstellenabschluss" und „erfolgreiche berufliche Integration". Somit können die hier betrachteten Kriterien aus wissenschaftlicher Sicht als sinnvolle Operationalisierungen des Lehrerfolgs bezeichnet werden. Schliesslich erfüllt das Vorgehen auch die Forderung der psychologischen Forschung, bei der Analyse des Lehrerfolgs mehrere Zielgrössen zu verwenden, da sowohl die Berufsschulnoten wie die beiden anderen betrachteten Kriterien mängelbehaftet sind (Baron-Boldt 1989).

In Bezug auf das Kriterium Berufsschulnoten stellt sich allerdings das Problem, dass die Examina und Noten der Lehre teilweise den gleichen Verzerrungen unterworfen sind, wie die Volksschulnoten (Baron-Boldt 1989). Entsprechend stellen Häfeli und Schellenberg (2009) fest, dass der Lehrerfolg in der Berufsschule relativ stark von der einzelnen Lehrperson abhängt. Dem Problem des Messfehlers in der abhängigen Variable kann nicht vollends begegnet werden. Zum einen werden aber in fast allen Berechnungen Zeugnisdurchschnittsnoten gebildet, womit die lehrerbedingte Fehlerkomponente reduziert wird. Zudem wird der Referenzgruppenfehler, der sich aufgrund unterschiedlicher Notenmassstäbe in den Berufsschulen ergibt, statistisch kontrolliert. Schliesslich sind Schulnoten trotz der Messfehlerproblematik ein legitimes und zentrales Kriterium des Lehrerfolgs, zumal sie entscheidend für die Promotionsentscheidungen in der Berufsschule sind.

2.6 Hypothesen zum Signalwert von Multicheck und Schulnoten

Was bedeuten die obigen Überlegungen zum Signalgehalt des Schulzeugnisses und des Multichecks für die Fragestellung dieser Arbeit? Bei der Beantwortung dieser Frage ist zu berücksichtigen, dass die dargestellten Befunde jeweils zu verschiedenen Zeitpunkten in den Arbeitsmärkten verschiedener Branchen und Berufe gemacht wurden, weshalb sie im Rahmen des in dieser Fallstudie betrachteten Detailhandels-Lehrlingsmarkt nicht oder nur bedingt von Bedeutung sein könnten. Trotzdem lassen sich einige Hypothesen zum Signalgehalt des Multichecks und jenem des Volksschulzeugnisses in Bezug auf die betrachteten Zielgrössen formulieren.

So ist, was die Vorhersage der *Berufsschulnotendurchschnitts* anbetrifft, zu erwarten, dass sowohl Multicheck wie auch Volksschulnoten (und Schulniveau) einen prognostischen Beitrag leisten kön-

[24] Die Nichtberücksichtigung betrieblicher Leistungen hat mehrere Gründe, darunter rechtliche, forschungsökonomische und Gründe der Vergleichbarkeit (keine einheitlichen und objektiven Leistungsbeurteilungen der Lernenden). Schliesslich geben die Lehrlingsverantwortlichen der Migros auch an, dass sie den Multicheck nicht für die Beurteilung der betrieblichen Leistungsfähigkeit eines Bewerbers einsetzen.

nen. Der Aussagegehalt des Multichecks dürfte sich dabei vor allem aus der Messung der effektiven schulischen Kompetenzen und der kognitiven Fähigkeiten ergeben, da beide Grössen mit den Berufsschulnoten wohl positiv korreliert sind. Die Schulnoten hingegen könnten einerseits aufgrund der regelmässigen Leistungsmessung, die ihnen zugrunde liegt, aber vor allem wegen ihrer sozialen Dimension einen positiven Signalwert aufweisen. Es ist indes zu erwarten, dass mit der Zeit vermehrt die tatsächlichen Kompetenzen der Schüler zum Tragen kommen, weshalb der Signalgehalt der Noten abnimmt (Baron-Boldt 1989; Neuenschwander und Malti 2009; Trapmann et al. 2007). Da Schulnoten und Schulniveau zudem weniger fachliche als soziale Kompetenzen messen, besteht die Möglichkeit, dass sich der Signalgehalt von Schulzeugnis und Multicheck in einer multivariaten Analyse eher ergänzen als aufheben.[25] Damit dürfte der Multicheck zu einem für die Wirtschaft verbesserten Signalling schulischer Kompetenzen beitragen (Bishop 1994). Allerdings dürften die Effektgrössen der Prädiktoren Multicheck und Volksschulnoten und die durch die Modelle insgesamt aufgeklärte Varianz in der abhängigen Variable relativ klein sein, da Messfehler in den Berufsschulnoten vorliegen und nichtkognitive Fähigkeiten und eine ganze Reihe anderer Lehrerfolgskenngrössen nicht oder nur indirekt berücksichtigt werden.

In Bezug auf die anderen betrachteten Zielgrössen kommt den Schulnoten wohl wiederum – stärker als ihre Leistungs- und Intelligenzkomponente – ihre soziale Komponente zugute. Lernbereitschaft oder Pflichtbewusstsein, welche gute Schulnoten mitunter signalisieren, dürften direkt sowohl mit einer geringeren *Wahrscheinlichkeit unentschuldigter Absenzen*, wie auch mit einer kleineren *Wahrscheinlichkeit einer Lehrvertragsauflösung* einhergehen. Zudem könnte sich der Signalgehalt der Noten vergrössern, weil ihr Anpassungsfaktor wohl generell mit der Zufriedenheit in der Lehre und dem Willen der Lernenden, sich mit der Lehrsituation zu arrangieren, zusammenhängt.

In beiden Fällen dürfte der Multicheck einen kleineren Signalwert aufweisen. So könnte sein Signalgehalt in Bezug auf unentschuldigte Berufsschulabsenzen darunter leiden, dass schulische und kognitive Fähigkeiten nicht direkt, sondern überwiegend indirekt über die Zufriedenheit mit der schulischen und betrieblichen Situation Einfluss auf die Zielgrösse ausüben. Hinsichtlich der Wahrscheinlichkeit einer Lehrvertragsauflösung ist aus theoretischer Sicht ebenfalls nur ein indirekter Zusammenhang zu erwarten, da gerade effektive schulische Fähigkeiten in vielen Fällen nur über die schulische Lehrsituation mit der Lehrabbruchquote zusammenhängen. Da der Multicheck jedoch die „berufsbezogene Eignung“ eines Bewerbers ausweist, sollte ihm bezüglich Lehrvertragsauflösungen allein aufgrund dieser Diagnose ein positiver Signalgehalt zukommen.

Im Hinblick auf das methodische Vorgehen der Studie sind vor allem zwei Feststellungen zu machen: Erstens könnte eine multivariate Analyse ohne die Berücksichtigung des Geschlechts, des Alters, und der sozialen, kulturellen und räumlichen Herkunft möglicherweise Verzerrungen aufweisen. Dementsprechend sind diese Merkmale zu erheben und statistisch zu kontrollieren. Zweitens gilt es mögliche Messfehler in der abhängigen Variable in der Auswertung zu berücksichtigen.

[25] Die Argumentation folgt Hell et al. (2008, 52). Die Autoren meinen, dass fachspezifische Studierfähigkeitstests über eine „substanzielle eigenständige Validität [verfügen] und zudem in der Lage [sind], die Prognosekraft von Schulnoten durch eine gemeinsame Berücksichtigung […] zu steigern“.

3 DATENGRUNDLAGE UND METHODIK

3.1 DATENGRUNDLAGE

3.1.1 STICHPROBE

Der Entscheid, den Signalgehalt des „Multicheck Detailhandel und Service" zu untersuchen, hatte hauptsächlich zwei Gründe: Erstens ist der Test relevant in dem Sinne, als es eine grosse Anzahl Detailhandelslernender in der Schweiz gibt. Zweitens hat der Test in seiner Branche eine grosse Bedeutung erlangt – mittlerweile verlangen alle grossen schweizerischen Detailhandelsunternehmen das Einsenden des Testresultats bei der Lehrstellenbewerbung. So wurde der Test allein im letzten Jahr von 10'300 Jugendlichen absolviert. Im Frühjahr 2010 wurden deshalb die grossen Detailhandelsunternehmen der Deutschschweiz angefragt, an der Studie teilzunehmen. Da die Firma Multicheck eine ähnliche Analyse erst kürzlich durchgeführt hatte (Multicheck 2008; Multicheck 2010c), waren die Verantwortlichen der Coop aufgrund des mit der Studie verbundenen Aufwands nicht zu einer Teilnahme zu bewegen. Die Verantwortlichen bei Denner und Volg zeigten derweil kaum Interesse an der Studie. Schliesslich beteiligten sich von den fünf angefragten Migros-Genossenschaften die Migros Aare, Luzern und Zürich an der Studie.[26]

In Gesprächen mit den Lehrlingsverantwortlichen jener drei Genossenschaften wurden daraufhin das weitere Vorgehen und die rechtlichen Bedingungen der Datenherausgabe erarbeitet. Zur Sicherstellung einer genügend grossen Stichprobe wurde entschieden, die Berufsschulzeugnisse des ersten und dritten Lehrsemesters zu berücksichtigen. Die Korrelation der Erst- und vor allem der Drittsemester-Zeugnisnoten mit den Noten an der Lehrabschlussprüfung – dem wichtigsten Kriterium des theoretischen Lehrerfolgs – dürfte relativ gross sein, auch wenn Berufsschulleistungen der Lernenden im Verlauf der Lehre ändern können, wie die Lehrlingsbetreuer der Migros in Gesprächen betonten. Schliesslich sind frühe Berufsschulleistungen vor allem aus zwei Gründen von Interesse: Erstens sind sie relevant in Bezug auf Multicheck und Schulnoten – beide Selektionsinstrumente dienen in erster Linie dazu, frühe berufsschulische Leistungen vorherzusagen. Zweitens ist der Einfluss des Lehrbetriebs auf die frühen schulischen Lehrleistungen der Lehrlinge geringer, was die überbetriebliche Generalisierbarkeit der Befunde dieser Arbeit erhöht.

Durch dieses Vorgehen wurde auch festgelegt, dass alle Detailhandelsfachlernende der Migros Aare, Luzern und Zürich, welche ihre Lehre im August 2007 oder im August 2008 aufgenommen hatten, in die Studie einbezogen werden (vgl. die Infobox auf der nächsten Seite). Die Stichprobe umfasste damit ursprünglich insgesamt 375 Detailhandelslernende – 171 Lernende der Migros Aare, 94 Lernende der Migros Luzern und 110 Lernende der Migros Zürich. Verschiedene Gründe führten allerdings zu einer Verkleinerung der Stichprobe. So wurden einerseits die so genannten zweimal zwei Lernenden nicht erfasst. Es handelt sich dabei um ungefähr 30 Lernende der drei Genossenschaften,

[26] Die Migros Basel verzichtete angesichts der geringen Anzahl an Detailfachlernenden auf eine Teilnahme, während die Migros Ostschweiz aufgrund einer Umstellung in der Datenverwaltung der Lernenden keine freien zeitlichen Ressourcen für eine Beteiligung an der Studie hatte.

welche zunächst eine zweijährige Lehre als Detailhandelsassistenten aufnahmen, und sich nun in zwei weiteren Jahren zum Detailhandelsfachmann beziehungsweise -fachfrau ausbilden.[27]
Zweitens erwiesen sich die Unterlagen in einigen Personaldossiers als unvollständig oder unbrauchbar. So fehlten beispielsweise Multicheck-Auswertungen, gewisse Schulzeugnisse, die Schulzeugnisse waren nur aus berufsvorbereitenden Schulen vorhanden oder stammten aus dem Ausland. Drittens wurden drei Lernende ausgeschlossen, welche lehrbegleitend die Berufsmittelschule besuchen.[28] Insgesamt ergibt sich somit ein Stichprobenumfang von 334 Lernenden. Da nicht all diese Lernende Berufsschulzeugnisse aus dem dritten Semester aufweisen, zum Beispiel aufgrund einer Lehrvertragsauflösung, verkleinert sich die Stichprobe auf 324 Lernende, wenn die Drittsemesternoten analysiert werden.

DIE MIGROS UND DIE DETAILHANDELSLEHRE IM ÜBERBLICK

Die Migros ist das grösste Unternehmen im Schweizer Detailhandel und zudem mit über 80‘000 Mitarbeitern auch der grösste Arbeitgeber des Landes. Pro Jahr werden mehr als 1‘000 Lehrstellen besetzt. Das Angebot der Migros-Gruppe umfasst die Bereiche Detailhandel, Industrie und Grosshandel, Kultur und Bildung, Freizeit und Sport, Gastronomie sowie weitere Bereiche wie Bank- und Druckereidienstleistungen. Sie ist in zehn Genossenschaften organisiert, von denen sich an der Studie die Migros Aare, Luzern und Zürich beteiligten. Die betrachteten Genossenschaften operieren im grössten Teil des Mittellandes (Bern, Solothurn, Aargau), in der Innerschweiz (Glarus, Luzern, Ob-/Nidwalden, Uri, Schwyz und Zug), in grossen Teilen des Kantons Zürich und in angrenzenden Orten des Kantons St. Gallen.
Die Migros erhält für die Lehrstellen im Detailhandelsfach je nach Ausbildungsberuf (z. B. Nahrungs- und Genussmittel, Textil, Fleischwirtschaft) und Migros-Region im Schnitt zwischen zehn- bis zwanzigmal mehr Bewerbungen, als sie Stellen zu vergeben hat. Die Auswahl der Lernenden erfolgt anschliessend in einem mehrstufigen Bewerbungsverfahren mit Bewerbungsschreiben, Schnupperwoche und Vorstellungsgespräch. So werden in den betrachteten Genossenschaften zurzeit pro Jahr ungefähr 200 Lernende zur Detailhandelsfachfrau beziehungsweise zum Detailhandelsfachmann ausgebildet. Dabei handelt sich um eine dreijährige Berufslehre, die zu einem Eidgenössischen Fähigkeitszeugnis (EFZ) führt. Im Schnitt arbeiten Detailhandelslernende während sieben Halbtagen die Woche im Betrieb, besuchen also während eineinhalb Tagen pro Woche die Berufsschule. Unterrichtet werden die Fächer Detailhandelskenntnisse, Deutsch, Französisch oder Englisch (nach Wahl), Gesellschaft, Wirtschaft und allgemeine Branchenkunde (nur während eines Semesters). Der Lehrplan wird durch überbetriebliche Kurse ergänzt und ist schweizweit einheitlich (BBT 2004).

[27] Auf die Berücksichtigung dieser Lernenden wurde verzichtet, weil sie offensichtlich die schulischen Erwartungen, welche ursprünglich an sie gerichtet wurden, übertreffen konnten. Mit anderen Worten handelt es sich um eine Gruppe von Lehrlingen, für welche der Lehrverlauf überdurchschnittlich positiv verlief. Somit hätte ihre Berücksichtigung möglicherweise einen verzerrenden Effekt auf die Ergebnisse gehabt.
[28] Die Berufsmittelschule im Detailhandel ist stark auf kaufmännische Bedürfnisse ausgelegt und hat einen mit der Detailhandelsfachlehre nicht vergleichbaren Lehrplan. Der sehr kleine Anteil an Lernenden, welche im Detailhandel die Berufsmatura lehrbegleitend absolvieren, entspricht der Quote in der Erhebung von Mühlemann et al. (2007).

3.1.2 GRUNDGESAMTHEIT UND DAS PROBLEM DER VORSELEKTION

Die Grundgesamtheit der Studie ergibt sich aus der Fragestellung. Es sind nicht nur diejenigen Jugendlichen, welche 2007 und 2008 bei der Migros tatsächlich eine Detailhandelslehrstelle erhielten. Im Grunde genommen interessiert auch der Signalgehalt der Multicheck-Ergebnisse und Schulnoten jener Volksschulabgänger, welche anschliessend keine Lehrstelle erhalten haben. Offensichtlich sind aber die Erfolgsvariablen für jene Jugendlichen nicht vorhanden, da sie nicht in die Berufslehre selektioniert wurden. Konkret äussert sich das Problem darin, dass die Schulnoten und das Multicheck-Ergebnis bei der Lehrlingsauswahl der Migros eine Rolle spielen und damit der unterste Teil der tatsächlichen Verteilung der beiden Selektionsinstrumente in den Bewerbungen nicht in der Stichprobe auftauchen. Das Problem manifestiert sich beim Multicheck stärker als bei den Schulnoten, enthält die Stichprobe doch 38 Lernende, welche sich mit einem Volksschulnotenschnitt beworben haben, der über beide erfassten Zeugnisse unter einer 4.0 liegt (vgl. Abbildung 5).[29] Sehr tiefe beziehungsweise ungenügende Multicheck-Gesamtergebnisse (unter 40 Punkten) sind in der Stichprobe hingegen fast keine vorhanden. Abbildung 1 macht die „Linkszensiertheit" der Variable mithilfe von simulierten Kernel-Dichtefunktionen deutlich (vgl. hierzu Cameron und Trivedi 2009). Sie zeigt einerseits die geschätzte Verteilung der Gesamtergebnisse der 7246 Detailhandels-Multichecks, die mit der 2006er Version des Tests (eingesetzt im Schuljahr 2006/2007) absolviert wurden (blau).[30] Diese streuen um den geeichten Mittelwert von 50 Punkten einigermassen symmetrisch. Die Verteilung der Gesamtergebnisse der Stichprobe dieser Arbeit ist indessen rechtsschief, Werte zwischen 50 und 65 Punkten sind die Regel und Werte unter 45 Punkten die klare Ausnahme (rot).

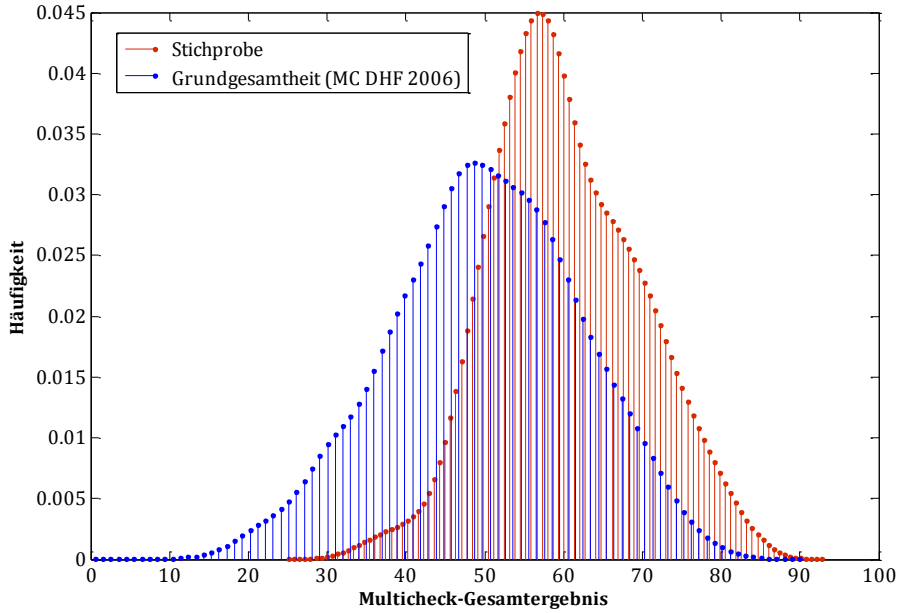
Die Darstellung illustriert zum Einen die Wichtigkeit des Multichecks in der Lehrlingsselektion. Nur 1.2% der Detailhandelsfach-Lehrlinge der Migros hatten sich mit einem ungenügenden Multicheck-Gesamtergebnis beworben. Dabei mag auch ein Selbstselektionseffekt eine Rolle gespielt haben: Personen, welche im Multicheck zu schlecht abschnitten, haben sich unter Umständen gar nicht mehr um eine Lehrstelle in diesem Bereich bemüht. Andererseits weist die Eingeschränktheit der Verteilung auf ein Problem dieser Arbeit hin: Der Signalgehalt des Multichecks kann nur über den oberen Teil seiner Verteilung beurteilt werden – ein Problem, das auch die bisherigen Studien zum Multicheck teilen (Multicheck 2008; Widmer 2006) und nur experimentell gelöst werden könnte. Es wird zu diskutieren sein, inwiefern dies die Schätzungen des Signalgehalts des Tests gegen unten verzerrt.

---

[29] Das Auftauchen solch tiefer Schulnoten in der Stichprobe hängt wohl einerseits damit zusammen, dass bei Jugendlichen, welche nach der Volksschule eine berufsvorbereitende Schule besuchten, vor allem das Zeugnis jener letzten Schule selektionsrelevant ist und eventuell besser ausfiel. Andererseits sind die schulischen Anforderungen des untersuchten Berufs eher gering.

[30] Die Daten wurden von der Firma Multicheck in Rohform zur Verfügung gestellt. Sie enthielten allerdings nicht die Gesamtergebnisse, wie sie auf den Auswertungen den Absolventen ausgewiesen wurden, sondern nur die Punkte der einzelnen Testgebiete und die dazu gehörenden Werte für den Zeit- und den Qualitätsbonus. (Der Algorithmus zur Errechnung des Gesamtergebnisses ist Geschäftsgeheimnis.) Somit mussten die Rohwerte selber zu einem Gesamtergebnis aggregiert werden. Dabei galt zu berücksichtigen, dass das Multicheck-Ergebnis gemäss den Aussagen der Verantwortlichen im Schnitt 50 Punkte beträgt. Deshalb wurde der über alle Testgebiete gebildete, gleichgewichtete Mittelwert von 57.1934 Punkten derart transformiert, dass er um die 50 Punkte streut. Das Vorgehen wurde von einem Verantwortlichen der Firma Multicheck überprüft und gutgeheissen.

ABBILDUNG 1: VERGLEICH DER VERTEILUNG DER IN DER STICHPROBE ERZIELTEN MULTICHECK-GESAMTERGEBNISSE (MCGESAMT, ROT) MIT DER GESCHÄTZTEN VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE DER GRUNDGESAMTHEIT (MULTICHECK VERSION 2006, BLAU)



3.1.3 *Datenerhebung und Erfassungsmethodik*

Die Datenerhebung erfolgte aus Datenschutzgründen vor Ort in den jeweiligen Personalabteilungen der Migros-Genossenschaften und dauerte neun Werktage. Die Daten wurden direkt aus den jeweiligen Personaldossiers erfasst. Bevor im nächsten Kapitel auf die einzelnen Variablen und deren Konstruktion eingegangen wird, werden in diesem Abschnitt einige erfassungsmethodische Vorgehensweisen diskutiert.

Schulsysteme und *Schultypen*: Das schweizerische Schulsystem gliedert sich auf der Sek-I-Stufe je nach Kanton in zwei bis vier Anforderungsniveaus. Die Schultypen wurden der Vergleichbarkeit halber analog der Publikationen des Bundesamtes für Statistik (BFS) in drei Anforderungsniveaus überführt (siehe bspw. BFS 2005b): Grundanforderungen, erweiterte Ansprüche und hohe Ansprüche. Die Einteilung entspricht den umgangssprachlich in gewissen Regionen üblichen Begriffen Real-, Sekundarschule und Spezsek. Daher wurden die dichotomen Variablen *real*, *sek* und *gym* gebildet. Diese Codierung wurde für alle Kantone verfolgt. Bei kooperativen und integrierten Schulsystemen entschied jeweils die Stammklasse des Schülers über die Zuteilung in die Kategorien.[31]

Berücksichtigte Schulzeugnisse: Es wurde entschieden, jeweils die beiden letzten Volksschulzeugnisse der Bewerbungsunterlagen zu berücksichtigen. Dabei bleibt mitunter unberücksichtigt, dass sich die Lernenden mit Schulzeugnissen der achten, neunten oder der zehnten Klasse beworben haben und sich der Signalwert der Zeugnisse möglicherweise nach Schuljahr systematisch unterscheidet. Das Vorgehen legitimiert sich aber aus dem Selektionsverfahren der Unternehmen, die ebenfalls die

[31] Zwar wurde bei der Erfassung vermerkt, welche Schulnoten auf einem höheren beziehungsweise tieferen Anforderungsniveau erreicht wurden. Es zeigt sich jedoch, dass der Signalwert unabhängig von der Schulstufe ist (vgl. Kapitel 4.4.2), so dass eine differenziertere Betrachtung jener Schulsysteme im Rahmen dieser Arbeit nicht angezeigt ist.

neuesten Zeugnisse einsetzen, um die Bewerberinnen und Bewerber für eine Lehrstelle auszuwählen (vgl. auch 4.2.2). Eine Ausnahme wurde bei denjenigen Lernenden gemacht, die sich mit Zeugnissen berufsvorbereitender Schulen (BVS) beworben hatten. Es zeigte sich nämlich, dass inhaltliche Ausrichtung, Schulfächer und auch Anforderungsniveau berufsvorbereitender Schulen nicht einfach aus den Zeugnissen ermittelbar sind. Deshalb wurde aus Gründen der Vergleichbarkeit, aber auch aufgrund der kleinen Anzahl jener Zeugnisse in den Unterlagen[32] und des Fokus' dieser Studie auf die Leistung der Volksschule auf die Erfassung der Zeugnisse berufsvorbereitender Schulen verzichtet. In solchen Fällen wurden stattdessen die neuesten Volksschulzeugnisse berücksichtigt.

Schulfächer und *Schulnoten*: Unterschiedlicher noch als die Schultypen sind die Lehrpläne der einzelnen Kantone. Sie unterscheiden sich nicht nur in den jeweils pro Fach aufgewendeten Schulstunden (BFS 2005a; SKBF 2010), sondern auch in den unterrichteten Fächern und deren Bezeichnungen. Deshalb wurde entschieden, nur der kleinste gemeinsame Nenner an Schulfächern zu erfassen. Es wurden die Noten der Landessprache Deutsch, der zwei Fremdsprachen Französisch und Englisch[33], die Noten in Mathematik[34], Naturwissenschaften[35] und in Geschichte und Geographie[36] erhoben.[37]

Berufsvorbereitende Schulen (BVS): 118 Lernende der Stichprobe haben den Berufseinstieg bei der Migros nicht direkt aus der Volksschule, sondern nach dem Umweg über eine berufsvorbereitende Schule geschafft. Wie oben angesprochen, gibt es in der Schweiz eine unüberschaubare Angebotsvielfalt berufsvorbereitender Schulen und die Ausrichtung und Zielsetzung dieser Schulen ist nicht einfach eruierbar (vgl. Imdorf 2005). Deshalb wurde die Variable *bvs* binär codiert. Sie nimmt den Wert eins an, wenn die Lernenden in den berufsvorbereitenden Schulen offensichtlich und in grösserem Umfang schulisches Wissen erarbeiteten, welches ihnen in Bezug auf die schulischen Anforderungen der Detailhandelslehre nützlich war. Dieses Vorgehen rechtfertigt sich dadurch, dass die Hauptfunktion von Zwischenjahren darin besteht, schulische Defizite aufzuholen (Herzog et al. 2004; Herzog et al. 2006).

---

[32] Da sich die Jugendlichen zum Zeitpunkt der Bewerbung oft noch in den berufsvorbereitenden Schuljahren befanden und meist noch keine Zwischen- oder Abschlusszeugnisse vorlagen, waren den Bewerbungsunterlagen nur wenige Zeugnisse jener Schulen beigelegt.

[33] In einigen Kantonen (z. B. Zürich) haben die Schüler allerdings die Möglichkeit, in der neunten Klasse nur noch eine Fremdsprache zu besuchen bzw. nur noch eine der beiden Fächer promotionsrelevant zu belegen. Deshalb sind nicht für alle Lernenden Volksschulnoten in beiden Fremdsprachen vorhanden.

[34] Noten in technischem Zeichnen wurden jeweils – zu einem Gewicht von einem Drittel – der Mathematiknote hinzugerechnet.

[35] Darunter werden insbesondere die Noten in Biologie, Physik und Chemie, aber auch die Noten des Faches NMM Natur verstanden.

[36] Die beiden Fächer wurden zusammengenommen, weil sie in einigen Kantonen in denselben Schulstunden unterrichtet werden. In Kantonen (bspw. Aargau), welche anstatt einzelner Fächer in den Natur- und Geisteswissenschaften nur das Fach „Realien" kennen, wurde die Realiennote sowohl der Naturwissenschafts- wie auch der Geschichts-/Geographienote zugerechnet – in den Notenschnitt allerdings nur einmal einbezogen.

[37] Auf die Erfassung der Noten der Bereiche Sport, Musik und Werken/Handarbeiten wurde verzichtet, weil der Bezug zu den berufsschulischen Anforderungen im Detailhandel nur gering ist.

3.2 Variablen und Transformationen

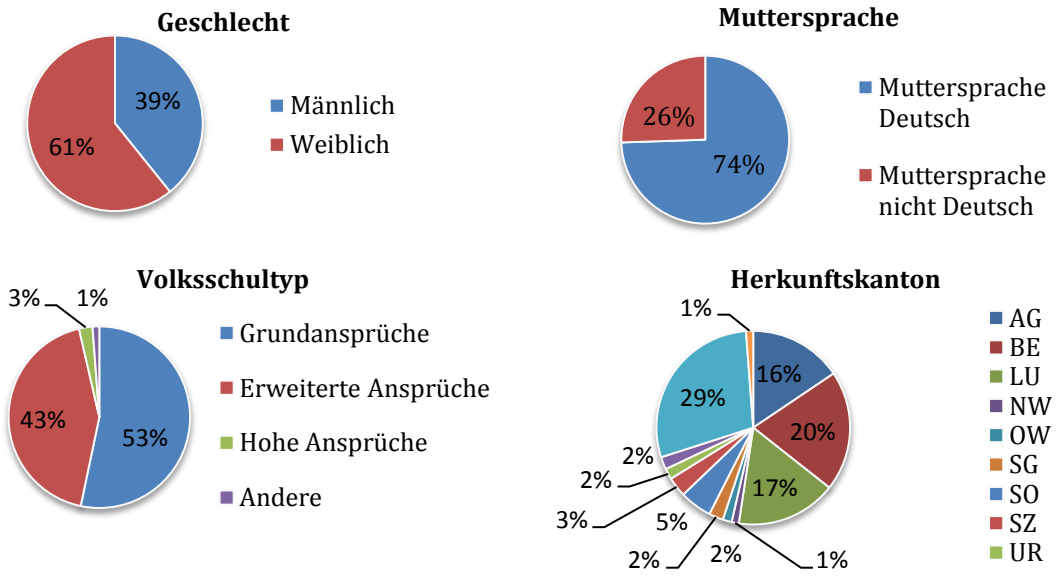
Tabelle 17 in Anhang 7.B zeigt, welche Variablen direkt aus den Personaldossiers erhoben wurden (kursiv jeweils die Stata-Bezeichnung der Variable). Aus den Berufsschulnoten wurde für beide Zeugnisse der Notenschnitt gebildet (*lschnitt1* bzw. *lschnitt3*). Dabei wurde die Note in den Detailhandelsfachkenntnissen – wie in der Lehrabschlussprüfung – doppelt gewichtet.

Die meisten anderen Variablen konnten telquel für die Analyse verwendet werden. Einige Variablen mussten allerdings transformiert oder umcodiert werden. So wurde erstens aus dem Geburtsdatum und dem Datum des Lehrbeginns das Lehrantrittsalter der Lernenden errechnet. Angesichts der Vermutung nichtlinearer Effekte des Alters – relativ alte und junge Lernende könnten gegenüber den Lernenden im Durchschnittsalter beispielsweise schlechtere Noten haben – wurden daraufhin zwei dichotome Variablen *laltera* und *lalterj* gebildet, welche den Wert eins annehmen, wenn ein Lernender zum obersten beziehungsweise untersten Quintil der Altersverteilung gehört.[38]

Um die räumliche Herkunft nicht nur in Bezug auf die Kantonszugehörigkeit zu berücksichtigen, wurde zweitens die Postleitzahl der besuchten Volksschule dazu verwendet, mithilfe der Raumgliederung der Schweiz eine Stadt-Land-Variable zu bilden. Die verwendete Stadt-Land-Typologie des BFS umfasst Werte von eins bis vier (vgl. BFS 2010). Die binäre Variable *stadt* nimmt den Wert eins an, wenn die Raumtypologie des Bundes einen Wert kleiner als zwei auswies.[39]

Schliesslich zeigten die Ausführungen in Kapitel zwei, dass der sozioökonomische Hintergrund der Lernenden möglicherweise einen Einfluss auf die Ergebnisse der Schätzungen hat. Zur Bestimmung des kulturellen Hintergrunds wurde deshalb aus den Lebensläufen die Muttersprache der Lernenden erhoben (*muttersp*). Als Proxy der sozialen Herkunft wurde die in den meisten Lebensläufen der Lernenden ausgewiesenen Berufe der Eltern erfasst. Zur Codierung bot sich die internationale Stan-

ABBILDUNG 2: STICHPROBENHÄUFIGKEITEN EINZELNER PERSÖNLICHER MERKMALE



[38] Der Grenzwert für *lalterj* liegt bei einem Alter von 16, für *laltera* bei einem Alter von 17.67 Jahren.

[39] Räume mit dem Wert eins sind städtische Gebiete, jene mit Wert zwei Agglomerationen und drei und vier bezeichnen ländliche Gebiete. Da die Postleitzahlen aus Datenschutzgründen nur bis auf drei Stellen genau erfasst werden durften, wurde jeweils der Mittelwert der Werte der Stadt-Land-Typologie all jener Ortschaften mit identischen ersten drei Ziffern gebildet und dieser dann zur Codierung der Variable *stadt* verwendet.

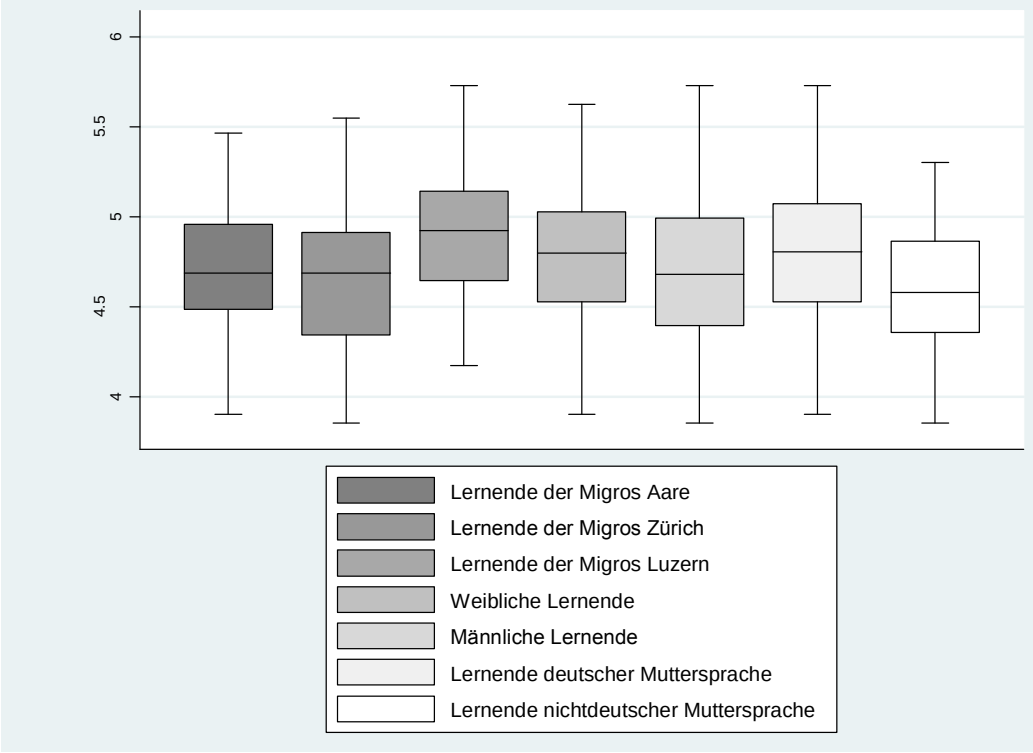
dardklassifikation der Berufe (ISCO) an.[40] Die Tabelle teilt die Berufe in zehn Hauptgruppen ein, welche hierarchisch nach vier Fähigkeitsstufen geordnet sind (ILO 2010).[41] Diesen vier Fähigkeitsstufen entsprechend wurde eine binäre Variable (*berufe12*) codiert. Sie nimmt den Wert eins an, wenn der Vater *oder* die Mutter einen Beruf ausübt, der nach ISCO der höchsten oder zweithöchsten Fähigkeiten-Kategorie zuzuordnen ist. Relevant war also jeweils die höhere Fähigkeitsstufe der beiden Berufe der Eltern.[42] Die so codierte Variable zeigt die erwarteten Zusammenhänge mit anderen Variablen: So korreliert sie beispielsweise signifikant mit der Muttersprache der Lernenden (ρ = 0.244).

3.3 Deskriptive Statistik

Abbildung 2 zeigt Stichprobenhäufigkeiten einiger zentraler Variablen. So sind 53 Prozent der Lernenden der Stichprobe Realschüler. Der Überhang der Frauen indes überrascht nicht, gilt das Detailhandelsfach doch als ein Frauenberuf (SKBF 2010; Moser 2004c).

Wie die Histogramme in Anhang 7.C zeigen, ist der Durchschnitt der beiden Bewerbungszeugnisse

ABBILDUNG 3: BOXPLOTS DER DURCHSCHNITTLICHEN BERUFSSCHULNOTE IM ERSTEN UND DRITTEN SEMESTER (LSCHNITT13) NACH MIGROS-GENOSSENSCHAFT, GESCHLECHT UND MUTTERSPRACHE



[40] Da die für die Schweiz adaptierte ISCO-Tabelle beim BFS nur kostenpflichtig bezogen werden kann, wurde die österreichische Version der Tabelle, ÖISCO-88 genannt, verwendet, welche auf der Internetseite der Statistik Austria gratis zur Verfügung gestellt wird. Die österreichische Klassifikation dürfte sich in den hier relevanten Inhalten – den Fähigkeitsstufen der Berufe – aufgrund der Ähnlichkeit des Bildungssystems und der Wirtschaftsstruktur kaum von der schweizerischen unterscheiden.

[41] Eine Alternative hätte darin bestanden, den *International Socio-Economic Index of Occupational Status* (ISEI) zu verwenden, welcher den sozioökonomischen Status der Berufe der Eltern auf einer Skala von 1 bis 100 einstuft (z. B. Kronig 2007). Allerdings sind die Berufsbezeichnungen in den Lebensläufen teilweise äusserst ungenau. So stellte sich die Frage, wie ein „Mitarbeiter der Firma XY" codiert werden sollte. Deshalb wurde auf eine spezifische und damit fehleranfälligere Erfassung des sozioökonomischen Status' verzichtet.

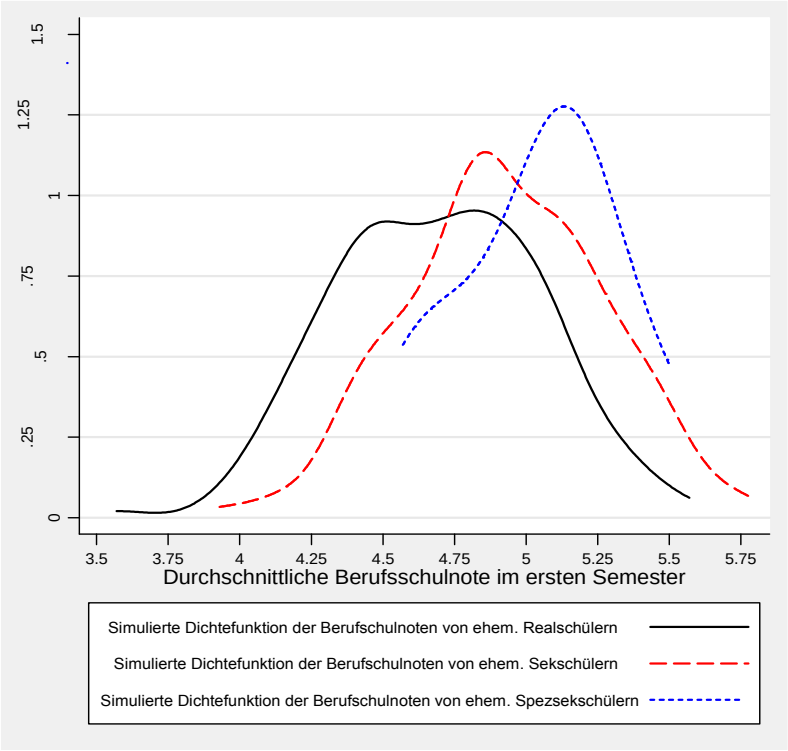
[42] Die Resultate der Schätzungen deuten darauf hin, dass nicht viel Information verloren geht, wenn die vier Kategorien des sozioökonomischen Hintergrunds zu zwei Kategorien zusammengefasst werden.

ziemlich exakt 4.5, während die Berufsschulnoten etwas höher sind. Die Standardabweichungen der Noten fallen dabei in allen vier betrachteten Schulzeugnissen in etwa gleich aus.

Die Boxplots in Abbildung 3 geben erste Hinweise über die Bedingtheiten der Berufsschulnoten. Zu beachten ist, dass die Noten der beiden Berufsschulzeugnisse zusammengenommen wurden (*lschnitt13*).[43] Es zeigt sich, dass die Lernenden der Migros Luzern bessere Durchschnittsnoten vorweisen können, als die Lernenden der anderen Regionen.[44] Unterschiede in den Noten weiblicher und männlicher Lernender sind kaum auszumachen. Im Gegensatz dazu haben Lernende nichtdeutscher Muttersprache tiefere Noten als Lernende deutscher Muttersprache.

Abbildung 4 illustriert schliesslich, dass die Berufsschulnoten eines ehemaligen Real- tiefer als jene eines einstigen Sekundarschülers sind. Die Abbildung verwendet dabei zur Illustration Verteilungsfunktionen der Erstsemester-Berufsschulnoten, welche mithilfe der Gauss-Kernfunktion simuliert wurden.[45] Konkret schlägt sich der Unterschied nach Schultyp in den Berufsschulnoten-Mittelwerten nieder: Während er für die ehemaligen Realschüler bei 4.68 liegt, ist er bei den ehemaligen Sekundarschülern mit 4.92 höher. Wie in der Abbildung ersichtlich ist die Streuung der Berufsschulnoten sehr ähnlich. Diese Befunde könnten darauf zurückzuführen sein, dass sich die auf Sek-I-Stufe vorhandenen (aber stark überlappenden) PISA-Leistungsunterschiede zwischen Real- und Sekundarschülern (Kronig 2007;BFS 2005b) in den Berufsschulnoten niederschlagen.

ABBILDUNG 4: SIMULIERTE DICHTEFUNKTIONEN DES BERUFSSCHULNOTENDURCHSCHNITTS IM ERSTEN SEMESTER (LSCHNITT1) NACH BESUCHTEM SCHULTYP IN DER VOLKSSCHULE



[43] Berücksichtigt wurden nur jene Lernenden, die in der Erfassungsperiode keinen Lehrabbruch verzeichneten, da ansonsten die Migros-Regionen nicht hätten verglichen werden können. Vgl. Kapitel 4.5.2.

[44] Dieses Ergebnis begründet sich wohl hauptsächlich mit den Unterschieden in der Notengebung der Berufsschulen. Vgl. Kapitel 4.2.

[45] Die unvollendete Form der Dichtefunktion für die ehemaligen Schüler einer Schule mit hohen Ansprüchen (*gym*=1) lässt sich auf die zu geringe Fallzahl zurückführen.

3.4 DETERMINANTEN DES DETAILHANDELS-MULTICHECKS

In diesem Abschnitt wird mithilfe linearer Regressionsmodelle des Multicheck-Gesamtergebnisses untersucht, inwiefern persönliche Merkmale und die Schulnoten der Lernenden mit dem Multicheck-Gesamtergebnis zusammenhängen. Dies wird im nächsten Kapitel dazu dienen, den Einfluss einzelner Variablen auf die geschätzten Koeffizienten des Multicheck-Gesamtergebnisses (*mcgesamt*) besser zu verstehen. Die Modelle erlauben es zudem, einige Fragen in Bezug auf die Testfairness und die Konstruktion des Tests anzugehen, die vor allem von Moser (2004c) aufgeworfen wurden.

Tabelle 1 zeigt die Ergebnisse der Schätzungen. Im ersten Modell werden neben dem Schnitt beider Volkschulzeugnisse (*zschnitt12*) Indikatoren wie das Alter bei der Multicheck-Testdurchführung (*mcalter*), der Volksschultyp (*sek, gym*), die Muttersprache und die Dummy-Variable *bvs* einbezogen. Diesen Kontrollvariablen werden im zweiten Modell die soziale Herkunft der Lernenden (*berufe12*), die Stadt-Land-Variable (*stadt*) sowie die binäre Variable, welche anzeigt, dass ein Absolvent den Multicheck zum zweiten Mal absolvierte (*mcrep*), angefügt. Das dritte Modell enthält schliesslich auch die Kantons-Dummy-Variablen.[46] Die Modelle vermögen zwar insgesamt wenig der Varianz in den Gesamtergebnissen des Multichecks zu erklären. Trotzdem zeigen sich interessante Muster.

So hängt der Mittelwert der Volksschulnoten (*zschnitt12*) nicht mit dem Multicheck-Gesamtergebnis zusammen. Dieses Resultat deckt sich mit den Ergebnissen in der psychologischen und pädagogischen Literatur, welche immer wieder eine tiefe Korrelation zwischen Eignungs- und Intelligenztests und Schulnoten festgestellt hat (Baron-Boldt 1989; Moser 2004c). Auch Widmer (2006) konstatierte einen geringen Zusammenhang zwischen KV-Multicheck und Schulnoten. Testregressionen mit dem Vollmodell zeigen zudem, dass einzig die über zwei Jahre gemittelte Mathematiknote (*zmath12*) auf einem Signifikanzniveau von 10% statistisch signifikant mit dem Multicheck-Gesamtergebnis zusammenhängt.

Zweitens schneiden Realschüler gegenüber den Sekundarschülern im Multicheck signifikant schlechter ab. Sollten Realschüler auf die Themen, die der Multicheck gemäss Aussagen des Herstellers entsprechend den Anforderungen in der Berufslehre erfasst, schlechter vorbereitet sein, ist dieser Effekt möglicherweise schlicht eine ungewollte praktische Konsequenz der inhaltlichen Ausrichtung des Tests. Unter Umständen bestätigt sich hiermit aber auch die von Moser (2004a, 47) geäusserte Vermutung, dass „Eignungstests […] Inhalte überprüfen, die dem Curriculum der Realschule weniger entsprechen“. Ob den Anforderungen entsprechend oder diskriminierend – mit den vorhandenen Daten ist nicht überprüfbar, wie gross die Benachteiligung der Realschüler im Multicheck tatsächlich ist, weil der Koeffizient des Schultyps möglicherweise aufgrund einer *Omitted Variable* verzerrt ist.[47] Trotzdem deutet auch die klare Benachteiligung der Jugendlichen aus dem Kanton Luzern darauf

[46] Die Referenzgruppe des Vollmodells sind männliche Berner Realschüler nichtdeutscher Muttersprache, die auf dem Land wohnen, keine BVS besuchten, Eltern mit tiefer sozialer Herkunft haben, und den Multicheck einmal absolvierten.

[47] Sollten Sekundarschüler nämlich höhere kognitive Fähigkeiten mitbringen als Realschüler, und diese kognitiven Fähigkeiten zudem zu höheren Ergebnissen im Detailhandels-Multicheck führen, dann würde der wahre Effekt des Schultyps in den Modellen in Tabelle 1 überschätzt. Diese Verzerrung wird trotz der Berücksichtigung der Volksschulnoten im obigen Modell, und im Übrigen auch trotz der Kontrolle mit dem PISA-Kurztest in Mosers‘ Studie, nicht (vollständig) korrigiert.

hin, dass der schulische Rucksack, den die Jugendlichen mitbringen, das Ergebnis im Test beeinflusst. Damit dürfte der Test die von einigen Autoren in ihn gesteckte Hoffnung wohl nur teilweise erfüllen, die systematische Benachteiligung der Realschüler im Übergang von Schule in die Berufsschule zu beheben (z. B. Häfeli und Schellenberg 2009) – denn *unabhängig* von der auf Sekundarstufe I besuchten Schulstufe sind die Ergebnisse im Detailhandels-Multicheck nicht.[48]

Die Modelle belegen weiter, dass Frauen im Detailhandels-Multicheck signifikant schlechter abschneiden als die männlichen Absolventen. Die Differenz zwischen den Geschlechtern macht sich auch in den 7246 Testergebnissen der 2006er Version des Detailhandelsmultichecks – den Resultaten der Grundgesamtheit – bemerkbar: Männer erzielen im Schnitt im selbst geschätzten[49] Gesamtresultat fünf Punkte mehr. Der Geschlechtseffekt, der auch schon von Widmer (2006) für den KV-

TABELLE 1: LINEARE REGRESSIONSMODELLE DES MULTICHECK-GESAMTERGEBNISSES

	mcgesamt		
	(1)	(2)	(3)
VARIABLEN	Basismodell	Erweitertes Modell	Vollmodell
zschnitt12	0.598	0.508	1.101
	(1.328)	(1.346)	(1.494)
weiblich	-2.837*	-2.312*	-2.232+
	(1.139)	(1.169)	(1.205)
muttersp	1.198	0.579	0.554
	(1.257)	(1.291)	(1.313)
mcalter	-1.104+	-1.157+	-0.748
	(0.607)	(0.607)	(0.628)
gym	10.81**	9.795**	10.48**
	(3.426)	(3.685)	(3.852)
sek	5.791**	5.793**	6.594**
	(1.148)	(1.148)	(1.295)
bvs	3.778**	3.654**	2.384+
	(1.302)	(1.298)	(1.358)
berufe12		2.605*	2.665*
		(1.167)	(1.171)
mcrep		0.0973	-0.0232
		(0.954)	(0.973)
stadt		0.816	0.517
		(1.512)	(1.550)
lu			-6.355**
			(1.905)
Konstante	71.72**	72.03**	65.52**
Observationen	249	249	249
Korr. R-Quadrat	0.132	0.140	0.150
AIC	1780	1781	1788
BIC	1808	1819	1865

Gegenüber der Referenzgruppe insignifikante Kantonsdummies nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[48] Hierzu schreibt Multicheck in den „häufig gestellten Fragen" auf ihrer Homepage: „Berücksichtigt Multicheck Junior die verschiedenen Schulstufen? Nein. Der Multicheck richtet sich nach den Anforderungen der verschiedenen Berufslehren. […] Damit wird den Lehrlingsverantwortlichen der direkte Vergleich der Leistungen ihrer Bewerberinnen und Bewerber ermöglicht."

[49] Vgl. Fussnote 30.

Multicheck konstatiert wurde, kann mehrere Gründe haben. So könnte er, da der Multicheck ein computerbasierter Test ist, auf die grössere Affinität männlicher Testpersonen zur Computertechnologie zurückzuführen sein (BFS 2005b). Andererseits könnte er entstehen, weil sich die Aspiranten für eine Detailhandelslehre tatsächlich je nach Geschlecht in ihrer Leistungsfähigkeit unterscheiden. Ein solcher Selbstselektionseffekt könnte etwa mit einem nach Geschlecht unterschiedlichen Berufsbild zu tun haben.[50]

Im Gegensatz zum Geschlechtseffekt, der ein Fairnessproblem des Multichecks andeutet, bewahrheitet sich eine andere Befürchtung, die gegenüber dem Verfahren geäussert wurde, nicht (Moser 2004c): Jugendliche, welche den Test erst relativ spät absolvieren, sind trotz ihres Wissensvorsprungs nicht signifikant besser als jene, welche ihn bereits relativ früh schreiben – der Koeffizient von *mcalter* ist insignifikant oder gar negativ. Dieses Resultat ist dabei unabhängig vom Einbezug der Variable *bvs*: Jugendliche, welche eine berufsvorbereitende Schule besuchten, erzielen ein höheres Multicheck-Gesamtergebnis unabhängig von der Tatsache, dass sie, wenn sie den Test ablegen, mehr als ein Jahr älter sind, als diejenigen, welche keine BVS besuchten.[51] Mit anderen Worten ist der Besuch der Schule und damit wohl ihr schulisches Vorwissen, aber nicht ihr Alter ausschlaggebend für den komparativen Vorteil. Schliesslich scheinen Jugendliche mit einer besseren sozialen Herkunft bessere Multicheck-Resultate zu erzielen. Dieses Resultat ist schwierig zu plausibilisieren, könnte aber beispielsweise mit einer besseren PC-Ausstattung und -Vertrautheit der Jugendlichen aus besseren Verhältnissen zusammenhängen.

[50] So könnte es sein, dass Männer die Detailhandelsfachlehre höher einschätzen als Frauen und dementsprechend die männlichen im Vergleich zu den weiblichen Jugendlichen in der Stichprobe höhere schulische oder kognitive Fähigkeiten aufweisen. Dies würde sich wiederum positiv auf das Multicheck-Ergebnis auswirken.

[51] Lernende, welche eine BVS besuchten, absolvierten den Multicheck im Schnitt mit 16.77 Jahren, während Lernende, die keine BVS absolvierten, den Test bereits mit 15.61 Jahren ablegten.

4 VORHERSAGEVALIDITÄTEN ALS SCHÄTZUNGEN DES SIGNALWERTS

In diesem Kapitel wird die im vorangehenden Kapitel dargestellte Datenbasis dazu verwendet, den Signalwert der Volksschulnoten mit jenem des standardisierten Berufseignungstests Multicheck zu vergleichen. Der Signalwert wird dabei anhand des prognostischen Gehalts der beiden Variablen in Bezug auf die drei erwähnten Zielgrössen eruiert. Die psychologische Forschung, aber auch die Sozial- und Wirtschaftswissenschaften sprechen in diesem Zusammenhang auch von einer Validitätsanalyse (Diekmann 2005; Eckstein 2008).[52] Spezifischer geht es darum, den Signalgehalt von Volksschulnoten und Multicheck anhand ihrer „prädiktiven Validität" oder „Vorhersagevalidität" abzuschätzen. Die Gliederung des Kapitels richtet sich nach dem methodischen Vorgehen und den betrachteten Kenngrössen des Lehrerfolgs. Zuerst steht der Zusammenhang der beiden betrachteten Instrumente mit den Berufsschulnoten des ersten und dritten Lehrsemesters im Fokus – zunächst in einem univariaten (Kapitel 4.1), anschliessend in einem multivariaten Kontext (Abschnitte 4.2 bis 4.4). In Abschnitt 4.5 wird schliesslich analysiert, inwiefern der Multicheck und die Volksschulnoten die Wahrscheinlichkeit von unentschuldigten Berufsschulabsenzen[53] und einer Lehrvertragsauflösung signalisieren.

4.1 KORRELATIONEN

Tabelle 2 zeigt Kreuzkorrelationen der fünf zentralen Untersuchungsvariablen. Erwartungsgemäss korrelieren die Noten der beiden Bewerbungszeugnisse (*zschnitt1* und *zschnitt2*) wie auch die beiden Lehrzeugnisse (*lschnitt1* und *lschnitt3*) stark. Ausserdem bestätigt sich wiederum, dass Multicheck-Gesamtergebnis (*mcgesamt*) und Volksschulnoten kaum zusammenhängen. Im Fall der Noten des zweitletzten Bewerbungszeugnisses ist die Korrelation sogar negativ signifikant. Es ist aber anzumerken, dass gemäss den Aussagen der Lehrlingsausbildner der Migros in der Selektionsentschei-

TABELLE 2: KREUZKORRELATIONEN DER DURCHSCHNITTLICHEN VOLKSSCHULNOTEN, DES MULTICHECK-GESAMTERGEBNISSES UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN

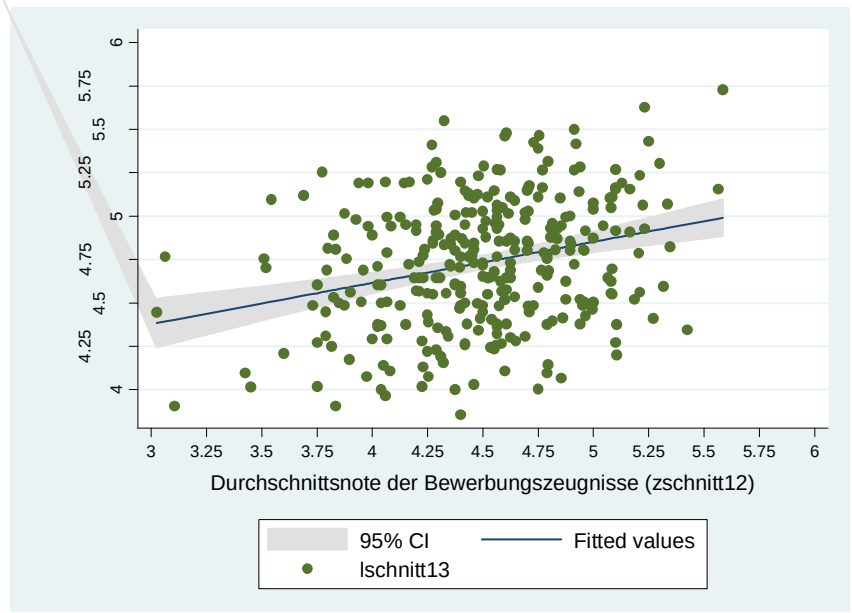
	zschnitt1	*zschnitt2*	*mcgesamt*	*lschnitt1*	*lschnitt3*
Zeugnisdurchschnitt letztes Bewerbungszeugnis *(zschnitt1)*	**1** n=334				
Zeugnisdurchschnitt zweitletztes Bewerbungszeugnis *(zschnitt2)*	**0.8116**** n=326	**1** n=326			
Gesamtergebnis Multicheck *(mcgesamt)*	**-0.0108** n=334	**-0.0939*** n=326	**1** n=334		
Zeugnisdurchschnitt 1. Semester Berufsschule *(lschnitt1)*	**0.2747**** n=331	**0.2195**** n=323	**0.1134**** n=331	**1** n=331	
Zeugnisdurchschnitt 3. Semester Berufsschule *(lschnitt3)*	**0.2905**** n=324	**0.2229**** n=316	**0.1192**** n=324	**0.737**** n=323	**1** n=324

** p<0.01, * p<0.05

[52] Es stellt sich die Frage, inwiefern der Begriff der Validität in Bezug auf Volksschulnoten angebracht ist, da sie nicht nur die Funktion haben, Berufsschulnoten zu prognostizieren. Trotzdem dürfte es letztlich eine der zentralen Aufgaben der Schulnoten der Sek-I-Stufe sein, Informationen über die Berufseignung eines Schülers zu liefern. Deshalb wird der Begriff Vorhersagevalidität in dieser Arbeit auch bei den Schulnoten verwendet.

[53] In diesem Zusammenhang kann nicht von Validität geredet werden, da sowohl Schulnoten wie auch Multicheck nicht vorgeben, unentschuldigte Absenzen vorherzusagen.

ABBILDUNG 5: TATSÄCHLICHE UND GEFITTETE WERTE EINER BINÄREN REGRESSION DER DURCHSCHNITTSNOTE DER BEIDEN LEHRZEUGNISSE AUF DIE DURCHSCHNITTSNOTE DER BEIDEN BEWERBUNGSZEUGNISSE



dung ein gutes Multicheck-Ergebnis schlechte Schulnoten und, umgekehrt, gute Schulnoten ein schlechtes Multicheck-Ergebnis kompensieren können, weshalb die Korrelation in der hier betrachteten Stichprobe geringer ausfallen könnte, als sie in Wahrheit ist. Trotzdem scheint sich die Hypothese zu bestätigen, dass Multicheck und Volksschulnoten eine andere Grösse messen, was im multivariaten Kontext dazu führen könnte, dass sich die Signalwerte der beiden Selektionshilfsmittel eher ergänzen als ausschliessen.

Die Tabelle gibt zudem bereits einen ersten Aufschluss über den Signalgehalt der Schulnoten und des Multichecks. Die beiden Zeugnisdurchschnitte der Volksschule hängen positiv und signifikant mit den Berufsschulzeugnissen zusammen – erwartungsgemäss schneidet dabei das neuere Zeugnis (*zschnitt1*) etwas besser ab. Die gefundenen Korrelationskoeffizienten liegen dabei etwas unter denjenigen, die Baron-Boldt (1989) in ihrer Metaanalyse älterer psychologischer Studien für den Zusammenhang zwischen Haupt- und Berufsschulnoten in Deutschland ausweist.[54] Der gefundene binäre Zusammenhang lässt sich am besten anhand der gemeinsamen Verteilung des Mittelwerts der Berufsschulnoten beider erfasster Semester (*lschnitt13*) und der durchschnittlichen Noten beider Bewerbungszeugnisse (*zschnitt12*) illustrieren (vgl. Abbildung 5).[55] Ebenfalls eingezeichnet wurde eine Regressionsgerade der binären Regression von *lschnitt13* auf *zschnitt12*.

Obwohl signifikant, ist der Korrelationskoeffizient des Multicheck-Gesamtergebnisses mit den Berufsschulnoten enttäuschend tief. Die Korrelation von knapp 0.12 liegt unter derjenigen, die Widmer (2006) für den kaufmännischen Multicheck findet, und deutlich unter der Korrelation von 0.4, welche die Firma Multicheck in ihrer eigenen Auswertung für das Gesamtergebnis des Detailhandels-Test ausweist (Multicheck 2008; Multicheck 2010c).[56] Aufgrund der fehlenden Dokumentation der

[54] Im Mittel findet Baron-Boldt eine (unkorrigierte) Korrelation von 0.461.

[55] War nur ein Bewerbungszeugnis vorhanden, wird in *zschnitt12* nur ein Zeugnisdurchschnitt berücksichtigt.

[56] Zu beachten ist, dass in Validitätsanalysen Korrelationen oftmals korrigiert werden. Insbesondere werden die Varianzeinschränkung der erklärenden – beispielsweise des Multichecks – sowie Ungenauigkeiten der

Multicheck-Studie ist nicht nachvollziehbar, wie der grosse Unterschied der Ergebnisse zustande kommt.

Die in Tabelle 3 aufgeführten Korrelationskoeffizienten zwischen den einzelnen Test*gebieten* des Multichecks und den Durchschnittsnoten der Lehrzeugnisse fallen ähnlich gering aus. Obwohl fünf der sieben Testgebiete jeweils mindestens einmal als signifikante Prädiktoren taugen, sind die gemessenen Effektgrössen gering. Der Konzentrations- wie auch der Logikteil des Tests haben keine Vorhersagevalidität. Generell lässt sich festhalten, dass diejenigen Teile des Tests, in denen Schulwissen abgefragt wird, valider sind. Entsprechend das übergeordnete Ergebnis im Bereich „Schulwissen" in Bezug auf die Berufsschulnoten aussagekräftiger, als das Ergebnis im Bereich „Potenzial".

4.2 Lineare Regressionsmodelle der Berufsschulnoten

In den folgenden Abschnitten wird der Signalgehalt des Multichecks und jene der Schulnoten in Bezug auf die Vorhersage von Berufsschulnoten innerhalb linearer Modelle (Abschnitt 4.2) und anschliessend aufgrund der spezifischen Form der abhängigen Variable innerhalb nichtlinearer Wahrscheinlichkeitsmodelle analysiert. Bestätigt sich in den Modellen der bescheidene Signalgehalt, den der Detailhandels-Multicheck in den univariaten Validitätsanalysen offenbart hat?

4.2.1 Modell, Spezifikationen und Ergebnisse

Es ist ein Standardverfahren der älteren Bildungsforschung, die Determinanten von Schulnoten mithilfe einfacher linearer Regressionsmodelle zu untersuchen (Fielding et al. 2003). Dieses Vorgehen unterstellt, dass die Intervallskala, die der Notenskala zu Grunde liegt, gleichmässig ist und die unabhängigen Variablen, insbesondere das Multicheck-Gesamtergebnis und die Noten der beiden Bewerbungszeugnisse, in linearer Form mit der abhängigen Variable, den Berufsschulnoten, zusammenhängen. Einfache Schätzungen mittels *Ordinary* (OLS) und *Feasible Generalized Least Squares* (FGLS) wurden zwar für die Daten ebenfalls durchgeführt (siehe Anhang 7.D), doch zeigt sich, dass die Mittelwerte der Berufsschulnotendurchschnitte *lschnitt1* und *lschnitt3* systematisch und statis-

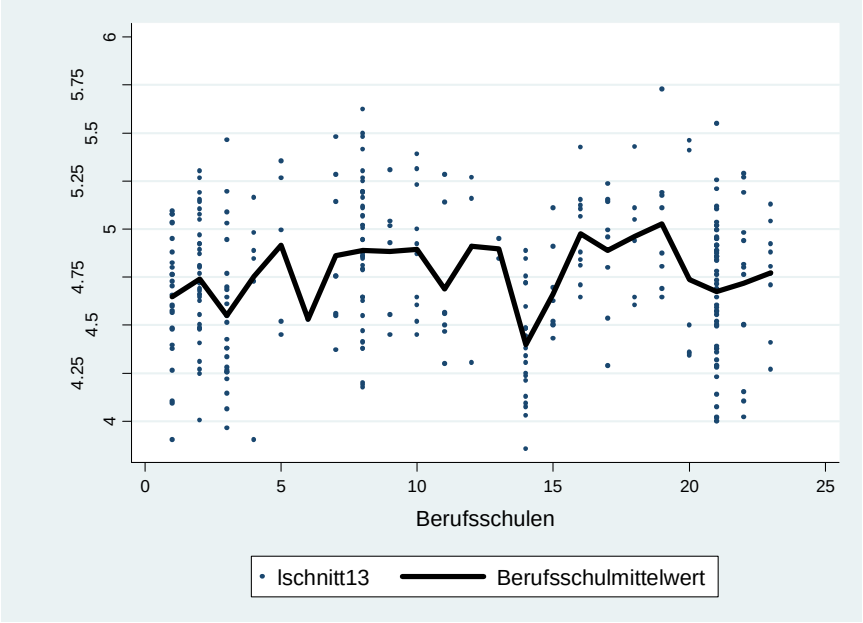
TABELLE 3: KREUZKORRELATIONEN DER MULTICHECK-TESTGEBIETE MIT DEN DURCHSCHNITTSNOTEN DER LEHRZEUGNISSE IM ERSTEN UND DRITTEN SEMESTER

	Multicheck-Testgebiet								
	Deutsch *(mcd)*	Franz. *(mcf)*	Englisch *(mce)*	Rechnen *(mcrechnen)*	Logik *(mclogik)*	Konzentration *(mckonz)*	Merkfähigkeit *(mcmerk)*	Potenzial *(mcpot)*	Schulwissen *(mcschul)*
Berufsschulnotendurchschnitt 1. Semester *(lschnitt1)*	**0.106*** n=331	**0.15**** n=209	**0.112*** n=297	**0.0969*** n=331	**-0.0386** n=331	**0.0437** n=331	**0.0918*** n=331	**0.0752** n=330	**0.1174**** n=330
Berufsschulnotendurchschnitt 3. Semester *(lschnitt3)*	**0.0392** n=324	**0.122*** n=203	**0.165**** n=290	**0.0840** n=324	**-0.0196** n=324	**0.0528** n=324	**0.117**** n=324	**0.0953*** n=323	**0.1116**** n=323

** p<0.01, * p<0.05

erklärten Variable statistisch berücksichtigt, weil beide Probleme zu einer Unterschätzung der tatsächlichen Validität der betrachteten Instrumente führen (Lienert und Raatz 1998). Die Berücksichtigung von Korrelationen zur Messung der Vorhersagevalidität steht aber in dieser Arbeit nicht im Zentrum und wird nicht weiterverfolgt.

ABBILDUNG 6: MITTELWERT UND STREUUNG DER BERUFSSCHULNOTEN DES ERSTEN UND DRITTEN SEMESTERS NACH BERUFSSCHULEN (N=23)



tisch signifikant und teilweise beachtlich nach Berufsschule variieren. Abbildung 7 visualisiert diese Tatsache.[57] Es bestätigt sich somit für unsere Daten die durch Amos et al. (2009) geäusserte Vermutung, dass der Schwierigkeitsgrad in den Berufsschulen trotz nationaler Leistungsstandards und äquivalenten Abschlüssen teilweise unterschiedlich ist.[58] Es ist damit angebracht, die Schätzungen innerhalb einer Modellstruktur durchzuführen, welche der Cluster-Struktur der abhängigen Variable Rechnung trägt und um die unbeobachtbaren Schuleffekte korrigiert.

Zur Schätzung geclusterter Stichproben greift die moderne Ökonometrie auf Panel-Daten-Methoden zurück. Die zwei populärsten Modelle, die in diesem Zusammenhang zur Anwendung gelangen, sind das *Random-Effects- (RE)* und das *Fixed-Effects-Modell (FE)* (Cameron und Trivedi 2009; Greene 2003; Schlotter et al. 2009; Wooldridge 2002a). Die Idee der Modelle besteht darin, die Mittelwertunterschiede (FE) der abhängigen Variable oder zumindest einen Teil davon (RE) mithilfe einer Mittelwertbereinigung auszugleichen. Die mathematische Formulierung des Modells, das im Folgenden geschätzt wird, macht deutlich, wieso diese Modelle auch als Mehrebenen- oder *Random-Intercept*-Modelle bezeichnet werden:

$$lschnitt_{ij} = \beta_1 + \beta_2 mcgesamt_{ij} + \beta_3 zschnitt12_{ij} + \boldsymbol{\beta}\mathbf{Z_{ij}} + \zeta_j + \varepsilon_{ij} \qquad (4.1)$$

[57] Es sei hier angemerkt, dass die teilweise wohl unterschiedlichen Notenmassstäbe der Berufsschulen aufgrund der identischen Funktion und Lehrpläne nicht mit Fairnessprinzipien zu vereinbaren wären, sofern sie nicht auf tatsächlichen Leistungsunterschieden der dortigen Lernenden beruhen. Dies ist aber bis zu einem gewissen Grad auszuschliessen, da (1) die Zuteilung zu den Berufsschulen zufällig ist bzw. aufgrund des Standortes der Ausbildungsstätte der Lernenden zustande kommt, (2) die Selektionskriterien für die Lehrstellen zumindest innerhalb der gleichen Migros-Genossenschaft ähnlich sind, und (3) auch einige Berufsschulen mit Lernenden desselben Herkunftskantons Mittelwertunterschiede aufweisen.

[58] Schätztechnisch stellt sich somit das Problem, dass in einer Schätzung mittels OLS der Erwartungswert eines Residuums gegeben die Kontrollvariablen, $E(\varepsilon_i|X)$, nicht unabhängig von einem Residuum ist, das von einer Person stammt, die in dieselbe Berufsschule geht – dies verletzt eine der Gauss-Markov-Annahmen.

Der Berufsschulnotenschnitt des *i*-ten Individuums in der *j*-ten Berufsschule hängt also von einer Linearkombination von Kovariaten ab. Dabei repräsentiert $\mathbf{Z_{ij}}$ einen Vektor von exogenen Kontrollvariablen. Der Einfachheit und Vergleichbarkeit halber wird dieser Vektor in der Folge für die meisten Spezifikationen gleich behalten, auch wenn einige der Koeffizienten in $\boldsymbol{\beta}$ teilweise insignifikant sind. Dies ist auch deshalb sinnvoll, weil vor allem die Koeffizienten β_2 und β_3 interessieren, und es damit jeweils weniger um die Modellauswahl und die statistische Effizienz der Schätzungen geht, als um die Robustheit der Ergebnisse in Bezug auf die beiden interessierenden Koeffizienten. Bei den in unserem Fall ausschliesslich dichotomen Kontrollvariablen in $\mathbf{Z_{ij}}$ handelt es sich um die Dummy-Variablen für das Alter (*laltera* und *lalterj*), das Geschlecht (*weiblich*), die Muttersprache (*muttersp*), die Raumtypologie der ehemaligen Schule der Lernenden (*stadt*) und den Dummy-Variablen für die verschiedenen Schultypen (*bvs, sek* und *gym*).

Die Modellspezifikation, der diese Kontrollvariablen zu Grunde liegen, wird in Bezug auf die verwendeten Variablen als „Basismodell" bezeichnet. Da sich in vielen Fällen regionale Unterschiede bemerkbar machen, werden die Kontrollvariablen des Basismodells jeweils im „erweiterten Basismodell" um die Dummy-Variablen der kantonalen Herkunft und um jene für die Migros-Genossenschaften (*luzern* und *zuerich*) erweitert.[59] Die Referenzgruppe im Basismodell sind damit jene männlichen Lernenden nichtdeutscher Muttersprache mittleren Alters, welche eine Realschule auf dem Land aber keine berufsvorbereitendes Schule besucht haben. Im erweiterten Basismodell stammen diese Lehrlinge zudem aus dem Kanton Aargau und haben eine Lehrstelle bei der Migros Aare.

Der Term ζ_j modelliert die Mittelwertunterschiede in der Notengebung der Berufsschule, indem er mit den Berufsschule *(j)* systematisch variiert. Der Parameter ist zur Modellauswahl entscheidend. Die effizientere RE- ist der FE-Regression vorzuziehen, wenn ζ_j als ein idiosynkratischer Fehlerterm mit Mittelwert null aufgefasst wird, der zudem mit allen Regressoren unkorreliert ist ($\mathrm{E}(\zeta_j|\mathrm{X}) = 0$). In unserem Fall sind allerdings beide Bedingungen nicht erfüllt: Einerseits ist nicht auszuschliessen, dass die Annahme $\mathrm{E}(\zeta_j|\mathrm{X}) = 0$ verletzt ist. Dies wäre zum Beispiel dann der Fall, wenn Berufsschulen deshalb tiefe Noten geben, weil es im jeweiligen Kanton üblich ist. Andererseits ist ζ_j keine zufällige Fehlerkomponente, sondern vielmehr als Parameter zu verstehen, den es analog einer Dummy-Variable zu schätzen gilt.[60] Deshalb werden nachfolgend Schul-Fixed-Effects-Modelle geschätzt.[61] (Wooldridge 2002b)

Das in (4.1) postulierte Modell unterstellt Homoskedastizität der Residuen ε_{ij}. Die in Abbildung 6 illustrierte Cluster-Struktur der abhängigen Variable legt allerdings die Vermutung nahe, dass auch die Varianz der Berufsschulnoten – analog dem Mittelwert – nach Berufsschule variiert. Konkret

[59] Gemäss den Informationskriterien und anderen Modellgüte-Kriterien ist teilweise das grössere, teilweise das kleinere Modell zu präferieren. Es wird deshalb nicht konsequent auf eines der beiden Modelle gesetzt, sondern in den meisten Fällen jeweils die Resultate beider Spezifikationen präsentiert.

[60] Ein wichtiger Nachteil des FE-Verfahrens – Variablen, die innerhalb des Clusters nicht variieren, können aufgrund der Mittelwertbereinigung nicht geschätzt werden – fällt im vorliegenden Fall nicht ins Gewicht, da keine solche Variablen vorliegen.

[61] Ein ähnliches Modell wie das in 4.1 skizzierte findet sich in BFS (2005b) oder Miller (1998).

führt dies in der Schätzung zu gruppenweise heteroskedastischen Residuen. Tests belegen diese Annahme zumindest in der FE-Regression für *lschnitt1*.[62] Es stellt sich somit die Frage, ob zur Inferenz des Modells ein Cluster-robuster Varianz-Kovarianz-Schätzer verwendet werden sollte. Die ökonometrische Literatur liefert keine einheitliche Antwort auf diese Frage, zumal die Eigenschaften des zur Anwendung kommenden, so genannten „Sandwich"-Schätzers gerade bei wenigen Clusters relativ schlecht untersucht sind (Wooldridge 2003). Das Problem besteht im Wesentlichen darin, dass der Schätzer in der Anzahl der Clusters – und nicht der Stichprobengrösse – konvergiert und sich dessen statistische Rechtfertigung aus seinen asymptotischen Eigenschaften ableitet. Deshalb tendiert der Cluster-robuste Schätzer in kleinen Stichproben mit wenigen Clustern dazu, zu konservativ zu sein (Cameron et al. 2008; Kézdi 2004).

Trotzdem wird im Folgenden – analog etwa zu Schwerdt et al. (2009) – auf den robusten Schätzer zurückgegriffen, da die mit dem normalen Schätzer berechneten Standardfehler der beiden interessierenden Koeffizienten jeweils nur geringfügig von jenen des robusten Schätzers abweichen.[63] Im Gegensatz dazu hängt die Signifikanz der Kantons- und Regionendummies stark von der Wahl des Schätzers ab. Da die Koeffizienten der Kantons-Dummy-Variablen ohnehin über alle Modelle hinweg stark variieren – die Ergebnisse diesbezüglich also alles andere als robust sind – wird auf eine Darstellung und Interpretation jener Koeffizienten meist verzichtet.[64]

Die Resultate der FE-Regressionen des Modells in (4.1) sind in Tabelle 4 ersichtlich. Die Ergebnisse sind dabei für die beiden erfassten Berufsschulzeugnisse (*lschnitt1* und *lschnitt3*) separat ausgewiesen. Dieses Vorgehen dient dazu, den zeitlichen Verlauf des Signalwerts zu berücksichtigen und die zum Schulnotensignalgehalt formulierte Hypothese zu überprüfen (vgl. Kapitel 2.6). Ausserdem werden in allen Modellen der Berufsschulnoten zwei Ausreisser weggelassen, welche jeweils einen erheblichen Hebeleffekt (*Leverage*) auf die geschätzten Koeffizienten, insbesondere auf jenen des Multicheck-Gesamtergebnisses, ausübten (vgl. Abbildung 7).

Entsprechend der in der Abbildung veranschaulichten, geringen Korrelation zwischen *mcgesamt* und *lschnitt1* ist der Koeffizient des Multicheck-Gesamtergebnisses in Tabelle 4 in den Modellen für die Berufsschulnoten des ersten Semesters nur auf dem 10%-Niveau signifikant. Die geschätzte Effektgrösse ist ausserdem bescheiden: Ein um 50 Punkte besseres Multicheck-Gesamtergebnis ergibt ceteris paribus einen um 0.212 Notenpunkte besseren Zeugnisdurchschnitt im ersten Semester der Berufsschule (Modell 1). Die Berufsschulnoten des dritten Semesters scheint der Multicheck nicht mehr vorherzusagen vermögen. Der Zeugnisdurchschnitt der Bewerbungszeugnisse – die Variable *zschnitt12* – erweist sich hingegen als valider Prädiktor beider Berufsschulzeugnisse. Lernende mit einem um eine Note besseren Bewerbungszeugnis erzielen im Schnitt ceteris paribus einen um etwa

[62] Sowohl der Lagrange-Multiplikator-Test von Greene (2003) wie auch das Testverfahren gemäss Baum (2006) verwerfen in den Modellen für *lschnitt1* die Nullhypothese der Homoskedastizität auf dem 10%-Signifikanzniveau. In den Modellen für *lschnitt3* ist keine Heteroskedastizität präsent.

[63] Eine elaborierte Lösung des Problems läge darin, bootstrapping-basierte Standardfehler zu schätzen (siehe Cameron et al. 2008).

[64] Zur Illustration werden in Tabelle 4 für die Koeffizienten von *mcgesamt, zschnitt12, luzern* und *zuerich* auch Standardabweichungen ausgewiesen, welche mit dem normalen Schätzer ermittelt wurden. Gerade für die beiden Regionen-Dummy-Variablen zeigen sich grosse Unterschiede.

0.29 Notenpunkte besseren Zeugnisdurchschnitt im ersten, und einen um 0.34 Notenpunkte besseren Zeugnisdurchschnitt im dritten Lehrsemester.[65]

Der Signalwert der Bewerbungs*zeugnisse* erhöht sich zudem durch die Signifikanz und Effektgrösse der Schultypusdummies. Gemäss dem Modell ist – bei sonst gleichen Vorbedingungen, insbesondere bei gleichen Volksschulnoten – ein Sekundarschüler um etwa 0.3 Notenpunkte besser als ein ehemaliger Realschüler. Anders herum ausgedrückt müssen aber die Noten eines Realschülers nur um ungefähr 0.3 Notenpunkte besser sein, als diejenigen eines Sekundarschülers, damit die erwarteten

TABELLE 4: FIXED-EFFECTS-REGRESSIONEN DER BERUFSSCHULNOTEN (BASISMODELL UND ERWEITERTES BASISMODELL)

	lschnitt1		*lschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweitertes Basismodell
mcgesamt	0.00424*	0.00373+	0.00233	0.00230
	(0.00207)	(0.00212)	(0.00255)	(0.00260)
	[0.00185]	[0.00184]	[0.00293]	[0.00305]
zschnitt12	0.298**	0.293**	0.340**	0.345**
	(0.0463)	(0.0472)	(0.0571)	(0.0581)
	[0.0435]	[0.0463]	[0.0529]	[0.0570]
laltera	0.0364	0.0327	-0.104+	-0.0945
	[0.0418]	[0.0416]	[0.0689]	[0.0682]
lalterj	0.0286	0.0124	0.0168	0.00281
	[0.0840]	[0.0865]	[0.0861]	[0.0859]
weiblich	0.0206	0.00836	0.0341	0.0173
	[0.0229]	[0.0287]	[0.0396]	[0.0440]
muttersp	0.140*	0.134*	0.0854*	0.0780+
	[0.0498]	[0.0517]	[0.0350]	[0.0402]
stadt	-0.0599	-0.0646	-0.142*	-0.141*
	[0.0502]	[0.0577]	[0.0511]	[0.0559]
bvs	0.0518	0.0550	0.0224	0.0319
	[0.0558]	[0.0594]	[0.0665]	[0.0729]
sek	0.286**	0.281**	0.307**	0.316**
	[0.0478]	[0.0411]	[0.0607]	[0.0613]
gym	0.486**	0.534**	0.658**	0.733**
	[0.115]	[0.109]	[0.122]	[0.125]
luzern		0.322**		0.185**
		(0.226)		(0.279)
		[0.0705]		[0.0589]
zuerich		-0.111		0.191
		(0.251)		(0.314)
		[0.163]		[0.115]
Konstante	2.942**	2.845**	2.796**	2.537**
Observationen	329	329	324	324
R-Quadrat	0.263	0.296	0.219	0.256
Cluster	23	23	23	23

Kantonsdummies nicht dargestellt
Cluster-robuste Standardfehler in eckigen Klammern, Standardfehler in runden Klammern
** p<0.01, * p<0.05, + p<0.1

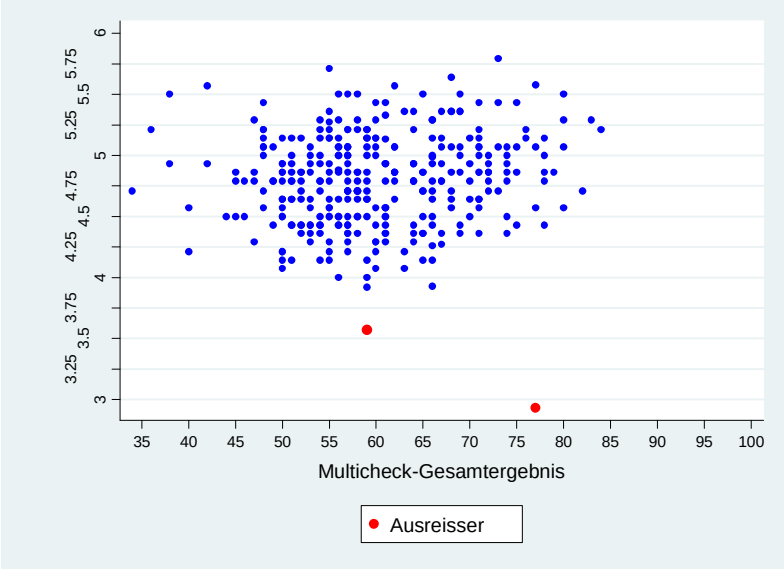
[65] Bemerkenswert ist dabei, dass die Vorhersagevalidität der Schulnoten – im Gegensatz zu jener des Multichecks – mit grösser werdender zeitlicher Distanz nicht abzunehmen scheint. Dies widerspricht der formulierten Hypothese, dass Schüler mit grösserer zeitlicher Distanz die Ungenauigkeiten der Notengebung je länger je mehr kompensieren können, und dementsprechend immer wie mehr ihre effektiven schulischen Fähigkeiten zum Tragen kommen.

Berufsschulnoten identisch sind. Dies bestätigt die grosse Leistungsüberlappung zwischen den beiden Schultypen, die sich auch in den PISA-Studien bemerkbar macht (BFS 2005b; SKBF 2010). Betrachtet man die weiteren Kontrollvariablen, fällt vor allem der Einfluss der Muttersprache auf, der sich in besseren Berufsschulleistungen niederschlägt. Eine städtische Herkunft scheint – ab dem zweiten Lehrjahr – hingegen mit schlechteren Ergebnissen in der Berufsschule einherzugehen. Alter und der Besuch einer berufsvorbereitenden Schule haben auf die erzielten Noten in der Berufsschule keinen wesentlichen Einfluss.

Das Modell lässt sich dazu verwenden, die zu erwartenden, durchschnittlichen Berufsschulnoten eines Lehrlings zu prognostizieren. Nimmt man das Basismodell für *lschnitt3*, ist bei einem Realschüler aus dem Kanton Aargau, der keine BVS besucht hat, deutscher Muttersprache ist, vom Land stammt, im Multicheck ein Gesamtergebnis von 60 Punkten erreichte und sich mit einem Zeugnisschnitt von 4.5 bewirbt, mit einer durchschnittlichen Drittsemester-Berufsschulnote von 4.55 zu rechnen. Hat derselbe Lernende stattdessen in der Sekundarschule eine 5.5 erreicht, ist eine 5.2 zu erwarten.

Zur Analyse des Signalgehalts des Multichecks und der Schulnoten im obigen FE-Modell können auch die Bestimmtheits- oder Gütemasse der Regressionen betrachtet werden. Es stellt sich in diesem Zusammenhang die Frage, wie gross der Beitrag der jeweiligen exogenen Variable zum erklärten Anteil der Varianz in den Berufsschulnoten ist.[66] Tabelle 5 zeigt die Gütemasse (R^2) der FE-Regressionen. Entfernt man den Volksschulnotendurchschnitt vom erweiterten Basismodell, sinkt der Anteil der durch das Modell erklärten Varianz in den Lehrnoten um 7 bis 8.5 Prozentpunkte. Wird das Multicheck-Gesamtergebnis ausgeklammert, sinkt das R-Quadrat nur marginal. Der Multicheck leistet also fast keinen und die Schulnoten nur einen kleinen Beitrag zur Erklärung der Differenzen in den Berufsschulnoten der Lernenden. Allerdings sind die Befunde auch in Relation zur

ABBILDUNG 7: GEMEINSAME VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE (MCGESAMT) UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM ERSTEN SEMESTER



[66] Siehe für eine Legitimation des Verfahrens Eckstein (2008).

gesamten Aussagekraft des erweiterten Basismodells zu betrachten. Hier zeigt sich, dass die schulischen Leistungsausweise – also Schultyp und Schulzeugnis – immerhin ähnlich viel Varianz in den Berufsschulnoten erklären, wie alle anderen berücksichtigen Variablen zusammen.[67]

TABELLE 5: BEITRAG DES MULTICHECK-GESAMTERGEBNISSES UND DER VOLKSSCHULNOTEN ZUM BESTIMMTHEITSMASS R^2 DER FIXED-EFFECTS-REGRESSIONEN DES ERWEITERTEN BASISMODELLS

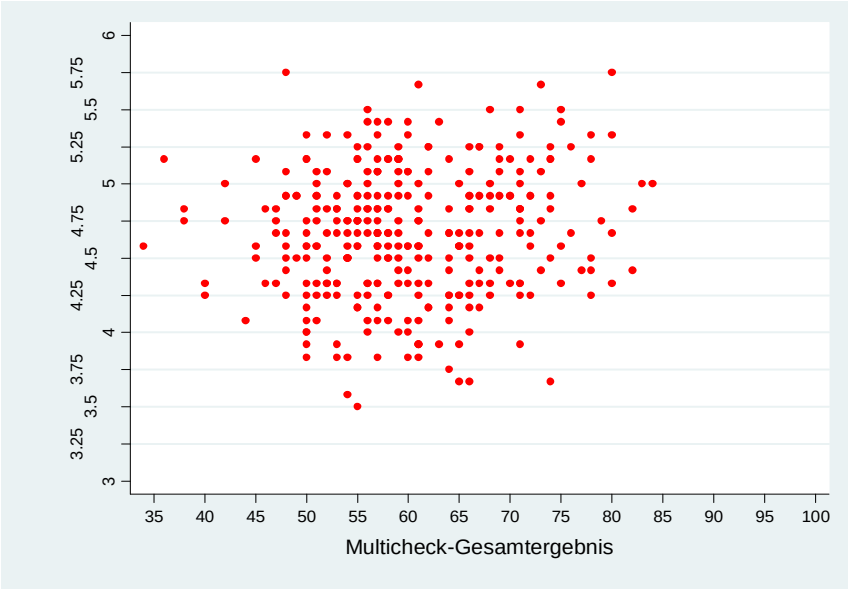
Modell	***zschnitt12* und *mcgesamt***	**Ohne *zschnitt12***	**Ohne *mcgesamt***
***lschnitt1* (Erweitertes Basismodell)**	$R^2 = 0.2940$	$R^2 = 0.2249$	$R^2 = 0.2851$
***lschnitt3* (Erweitertes Basismodell)**	$R^2 = 0.2415$	$R^2 = 0.1585$	$R^2 = 0.2402$

4.2.2 Robustheit

Im Folgenden soll die Robustheit der aus den FE-Modellen gewonnen Erkenntnisse überprüft und allenfalls qualifiziert werden. Wie bereits erwähnt, wurden das Modell in (4.1) – ohne die berufsschulspezifische Konstante – auch mittels OLS und GLS geschätzt (vgl. Anhang 7.D).[68] Die so ermittelten Koeffizienten sind bezüglich Signifikanz und Effektgrösse sehr ähnlich wie jene der FE-Modelle. Alle geschätzten linearen Regressionsmodelle liefern damit in Bezug auf den Signalwert der beiden zentralen Variablen ein konsistentes Bild.

Allerdings kommt bei der Betrachtung eines Streudiagramms, welches die erzielten Multicheck-Gesamtergebnisse den Berufsschulnoten gegenüberstellt, die Frage auf, ob die Validität des Multicheck-Gesamtergebnisses (*mcgesamt)* gegen oben hin zunimmt (siehe Abbildung 8). Der visuelle Eindruck bestätigt sich, wenn beim Mittelwert des Multicheck-Gesamtergebnisses (60.17) separat

ABBILDUNG 8: GEMEINSAME VERTEILUNG DER MULTICHECK-GESAMTERGEBNISSE (MCGESAMT) UND DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM DRITTEN SEMESTER



[67] Ohne *zschnitt12, sek* und *gym* sinkt R^2 im Modell für *lschnitt3* auf 0.1086, im Modell für *lschnitt1* auf 0.1788.
[68] Feasible GLS hat dabei gegenüber OLS den Vorteil, dass es die in den Modellen für *lschnitt1* vorliegende, gruppenweise Heteroskedastizität der Residuen berücksichtigt und in diesem Fall effizienter ist (Greene 2003).

berechnete Korrelationskoeffizienten zwischen *mcgesamt* und *lschnitt3* verglichen werden ($\rho_{mcgesamt\leq 60}$ = 0.052 gegenüber $\rho_{mcgesamt>60}$ = 0.281).[69] Ob dieser Befunde auch in einem multivariaten Kontext zum Tragen kommt, kann mithilfe einer sogenannten *Spline Regression* ermittelt werden (Greene 2003). Dabei wird die Regressionsgerade in eine stückweise lineare Funktion umgewandelt. In unserem Fall wird der Steigungsparameter des Multicheck-Gesamtergebnisses beim Mittelwert zweigeteilt und zwei Steigungsparameter sowie zwei Konstanten geschätzt – bildlich gesprochen erhält die Regressionsgerade in der Mitte einen Knick.[70]

Tabelle 6 enthält die Ergebnisse dieser Regression. Das Bild, das sich ergibt, ist weniger eindeutig, als in der binären Analyse: So ist der Koeffizient für *mcgesamt1* – für Multicheck-Gesamtergebnisse mit höchstens 60 Punkten – zwar jeweils insignifikant und sehr klein. Doch auch sein Pendant für Gesamtergebnisse ab 60 Punkten, *mcgesamt2*, ist jeweils nur im kleineren Modell und auf dem 10%-Niveau signifikant.[71] Dass die Aussagekraft des Multicheck-Gesamtergebnisses gegen oben hin zunimmt, kann in den Fixed-Effects-Modellen also nur teilweise gestützt werden.

In Bezug die Koeffizienten der Schulnoten stellt sich eine andere Frage: Ist die Validität der Schulnoten robust gegenüber der Wahl des Schuljahres, aus dem die Schulnoten stammen? Denn bislang wurde von der Richtigkeit ausgegangen, als erklärende Variable den Schulzeugnisschnitt der beiden aktuellsten Bewerbungszeugnisse (*zschnitt12*) einzusetzen (vgl. Kapitel 3.1.3). Es besteht allerdings

TABELLE 6: FIXED-EFFECTS-REGRESSIONEN DER BERUFSSCHULNOTEN MIT BEIM MITTELWERT GETEILTEM STEIGUNGSPARAMETER FÜR DAS MULTICHECK-GESAMTERGEBNIS (SPLINE REGRESSION)

	lschnitt1		*lschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweitertes Basismodell
mcgesamt1	0.000245	-0.000905	0.00853	0.00725
	(0.00498)	(0.00538)	(0.00521)	(0.00567)
mcgesamt2	0.00911+	0.00788	0.0127+	0.0124
	(0.00490)	(0.00470)	(0.00722)	(0.00736)
zschnitt12	0.296**	0.291**	0.336**	0.338**
	(0.0357)	(0.0326)	(0.0534)	(0.0566)
Observationen	329	329	324	324
R-Quadrat	0.268	0.300	0.237	0.271
Cluster	23	23	23	23

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[69] Die Teilung der Multicheck-Gesamtergebnisse bei 60 Punkten macht auch deshalb Sinn, weil Multicheck ein Gesamtergebnis von 60 Punkten und mehr als übertroffen, ein Ergebnis darunter als erfüllt bzw. nicht erfüllt einstuft.

[70] Wobei hier darauf verzichtet wird, die Parameter derart zu restringieren, dass eine kontinuierliche Regressionsgerade entsteht. Die Gerade weist beim Mittelwert von *mcgesamt* also einen kleinen Sprung auf.

[71] Ausserdem wurde überprüft, ob sich der Signalgehalt des Multicheck-Gesamtergebnisses allenfalls nach verwendeter Testversion oder nach Testdurchführungsjahr systematisch unterscheidet. Die Ergebnisse deuten jedoch nicht darauf hin. So sind die Koeffizienten der Multicheck-Gesamtergebnisse in der Testversion des Multichecks, die vom 1. Juli 2006 bis zum 31. Juni 2007 eingesetzt wurde, im Basismodell ebenso insignifikant, wie jene Koeffizienten der Multicheck-Gesamtergebnisse, die auf der vom 1. Juli 2007 bis zum 31. Juni 2008 eingesetzten Version beruhen (vgl. 7.E, Tabelle 20). Die Resultate entsprechen den Aussagen der Multicheck-Verantwortlichen, der Detailhandels-Multicheck habe seit 2006 keine grösseren Veränderungen erfahren.

durchaus Grund zur Annahme, dass der Signalwert der Schulnoten vom jeweiligen Schuljahr, aus dem die Schulnoten stammen, abhängt – beispielsweise weil sich jene Personen, die sich mit einem Zeugnis der neunten Klasse bewerben, in nicht beobachtbaren Eigenschaften systematisch von jenen unterscheiden könnten, die sich mit einem Zeugnis der achten Klasse bewerben. Die nach Semesterzeugnissen gesonderte Validität der Schulnoten lässt sich am einfachsten feststellen, wenn die Modelle einzeln für jedes Semesterzeugnis geschätzt werden. Die Ergebnisse dieser Regressionen (vgl. Anhang 7.E, Tabelle 21) deuten darauf hin, dass sich die Schulzeugnisse in ihrer prädiktiven Validität nicht wesentlich unterscheiden. So fällt einzig auf, dass die beiden Neuntklasszeugnisse (*zschnitt95* und *zschnitt9*) eine im Vergleich zu den Zeugnissen aus dem achten Schuljahr (*zschnitt85* und *zschnitt8*) etwas geringere Effektgrösse aufweisen, wenn es um die Vorhersage der Berufsschulnoten des ersten Semesters geht.[72] Die Ergebnisse bestätigen aber, dass die Schulnoten unabhängig davon, aus welchem Semester sie stammen, signifikant den Lehrerfolg vorherzusagen vermögen und sich die Höhe des gemessenen Effekts in einem relativ engen Band bewegt. Das gewählte Vorgehen, als Schulnotenvariable jeweils den Mittelwert aus den beiden aktuellsten Bewerbungszeugnissen zu verwenden, ist damit gerechtfertigt.

Schliesslich gilt es, die in der Einleitung zur FE-Regression erwähnte Annahme zu überprüfen, dass das Modell Normalität der Residuen und – damit verknüpft – eine gleichmässige Intervallskala unterstellt. Hierzu können Normalquantil-Plots der Residuen der Regressionen verwendet werden (vgl. Anhang 7.F). Es zeigt sich, dass die Residuen der erweiterten Basismodelle für *lschnitt1* und *lschnitt3* normalverteilt sind, da die Quantile der Residuen der theoretischen Quantile der Normalverteilung (Diagonale) folgen. Die anderen Modelle (OLS und GLS) und Spezifikationen weisen ähnlich normale Residuen auf. Damit lässt sich auch festhalten, dass die Modelle richtig spezifiziert sind und keine *nichtlinearen* Effekte ausser Acht gelassen wurden. Dies impliziert insbesondere auch, dass die Schweizer Notenskala die Annahme der Intervallskala der linearen Regression grösstenteils zu erfüllen scheint. Mit anderen Worten: Der zu erwartende Unterschied in den Berufsschulnoten zwischen zwei Bewerbern, die sich mit der Note 4 beziehungsweise 4.5 bewerben, ist identisch mit dem Notenunterschied, der zwischen Personen mit einer 5 und einer 5.5 zu erwarten ist.[73]

4.3 Nichtlineare Wahrscheinlichkeitsmodelle der Berufsschulnoten

4.3.1 Problemstellung und Modell

Zwar ist das Fixed-Effects-Modell korrekt spezifiziert und es scheint im vorliegenden Kontext gut zur Modellierung von Berufsschulnoten eingesetzt werden zu können. Doch hat es sich gezeigt, dass der geclusterte Charakter der abhängigen Variable die Schätzung wie auch die Inferenz in der linearen Regression beeinflusst. Die gewonnenen Erkenntnisse sollen daher im Rahmen von nichtlinearen

[72] Dies muss allerdings nicht zwingend als Evidenz dafür gewertet werden, dass die Zeugnisse der neunten Klasse in der Vorhersage früher Lehrleistungen weniger aussagekräftig sind. Vielmehr haben jene, die sich mit einem Zeugnis der neunten Klasse bewerben, wesentlich häufiger eine berufsvorbereitende Schule besucht. Deshalb ist die prädiktive Validität eher auf die grössere zeitliche Distanz zwischen Schul- und Lehrzeugnissen zurückzuführen. Umso bemerkenswerter ist es, dass die geschätzte Effektgrösse der Zeugnisse der neunten Klasse im Modell für *lschnitt3* tendenziell ansteigt.

[73] Dieser Eindruck entsteht auch bei der Betrachtung des Streudiagramms in Abbildung 5.

Wahrscheinlichkeitsmodellen für geordnete Kategorien überprüft und vertieft werden (Boes und Winkelmann 2006; Cameron und Trivedi 2009; Greene 2003; Wooldridge 2002a). Dies entspricht dem Ansatz der neueren bildungsökonomischen Forschung und ist gerade im anglo-amerikanischen Raum, welcher in den Zeugnissen Buchstaben statt Ziffern kennt, populär. Einige Autoren haben explizit darauf hingewiesen, dass die Verwendung von Modellen für geordnete Kategorien anderen Ansätzen wie der FE-Regression vorzuziehen ist, da es sich bei Noten zwar um hierarchische, aber nicht intervallskalierte Variablen handle (z. B. Grant 2007; Sund 2009). So weisen auch Fielding et al. (2003) auf die Vorteile dieser Modelle bei der Analyse von Schulnoten hin. Es sind dies vor allem die weniger starken Annahmen über die Skaliertheit der Noten und die geringere Clusterabhängigkeit des Verfahrens.

Zwar scheint es in unserem Fall durchaus angebracht, den Noten Normalität, Linearität und ein konstantes Intervall zu unterstellen. Wie oben angedeutet, besteht der Vorteil des Verfahrens im Kontext dieser Arbeit aber darin, dass sich bei entsprechender Kodierung der abhängigen Variable das Problem der nach Berufsschulen unterschiedlichen Notengebungen lösen lässt. Die Idee besteht darin, eine Variable mit drei Kategorien zu bilden, welche die Lehrlinge *relativ* zu den Leistungen der anderen Schüler *in der jeweiligen Berufsschule* einordnet. Deshalb wurde für die zwei erfassten Lehrzeugnisse je eine Variable gebildet, die den Wert null annimmt, wenn es sich beim betreffenden Lernenden um einen Schüler handelt, der sich im untersten Quartil der Notenverteilung der jeweiligen Berufsschule befindet; die den Wert zwei annimmt, wenn die Leistungen des Lernenden zum obersten Quartil der Notenverteilung gehören; oder die den Wert eins annimmt, wenn sich die Leistungen des Lernenden in den mittleren Quartilen befindet. Die so gebildeten abhängigen Variablen werden als *lquartillschnitt1* und *lquartillschnitt3* bezeichnet.[74]

Der Nachteil des Vorgehens ist, dass die Transformation der Noten in eine geordnete, nur drei Kategorien umfassende Variable mit einem Informationsverlust verbunden ist. Zwei Gründe legitimieren die Methode aber: Erstens sind bei den offensichtlich unterschiedlichen Ansätzen der Notengebung absolute Vergleiche der Notendurchschnitte von Schülern verschiedener Schulen und Lehrern sowieso nur bedingt zulässig. Zweitens dürfte die Variable, obwohl sie zugegebenermassen eher ein theoretisches Konstrukt ist, für die Lehrlingsausbildner insofern verwertbar sein, als sie die unterschiedlichen Notenniveaus der Berufsschulen teilweise kennen. Daher ist es für sie eine nützliche Information, ob ein Schüler mit gewissen Merkmalen zu den Besten, dem Durchschnitt oder den Schlechtesten in der jeweiligen Berufsschule gehören wird – sie können diese Information schliesslich auf eine ungefähre Notenskala übertragen.

Zur Identifizierung der Modelle für geordnete Kategorien wird angenommen, dass der abhängigen, kategorischen Variable eine unbeobachtbare latente Variable zu Grunde liegt (Wooldridge 2002a). Um die latenten, wahren Schülerleistungen (*lquartillschnitt*)* in die abhängige, beobachtbare Variab-

[74] Da die Bildung von Quartilen für Berufsschulen, in denen weniger als sechs Lernende zur Schule gehen, wenig Sinn macht, wurde für die Lernenden jener Berufsschulen eine Residualberufsschule kreiert. Anschliessend wurde ihre relative Position in der Verteilung dieser neu kreierten Berufsschule ermittelt.

le (*lquartillschnitt*) umzuformen, werden deshalb zusammen mit dem Modell zwei Grenzwerte (α_1 und α_2) geschätzt.[75] Es ergibt sich folgende Formulierung:

$$lquartillschnitt_i = \begin{cases} 1 \; if \; -\infty < lquartillschnitt_i^* < \alpha_1 \\ 2 \; if \; \alpha_1 < lquartillschnitt_i^* < \alpha_2 \\ 3 \; if \; \alpha_2 < lquartillschnitt_i^* < \infty \end{cases}$$

Daraus lässt sich ein Modell für die latente Variable ableiten:

$$lquartillschnitt_i^* = \beta_1 + \beta_2 mcgesamt_i + \beta_3 zschnitt12_i + \boldsymbol{\beta}\mathbf{Z_i} + \varepsilon_i \qquad (4.2)$$

Wird dem Residuum ε_i eine Verteilung unterstellt, lassen sich mithilfe der kumulativen Dichtefunktion Eintretenswahrscheinlichkeiten für *lquartillschnitt* berechnen (Cameron und Trivedi 2009). In der Praxis werden dabei oft zwei Verteilungen favorisiert: die logistische und die standardnormale. Diese begründen das sogenannte Ordered-Logit- beziehungsweise das Ordered-Probit-Modell. Die Entscheidung zwischen den beiden Modellen ist nicht trivial (Greene 2003). Im vorliegenden Fall können wir uns allerdings zu Nutze machen, dass eine Form des latenten Modelles in (4.2) bereits in den vorangehenden Abschnitten geschätzt wurde und damit die Verteilung der Residuen beobachtet werden konnte. Aufgrund der Normalität der Residuen, die in den Normalquantil-Plots (Anhang 7.F) zutage kam, werden deshalb nachfolgend Ordered-Probit-Modelle geschätzt. Allerdings wurden die Resultate einiger wichtiger Regressionen (beispielsweise jene der Tabelle 7) auch mit einer logistischen Verteilung geschätzt.[76] Die Resultate hängen dabei kaum von der Wahl der Verteilung ab, weshalb auf eine Darstellung der Ordered-Logit-Regressionen verzichtet wird.[77]

4.3.2 Ergebnisse

Die Ergebnisse der Ordered-Probit-Regressionen des Modells in (4.2) sind in Tabelle 7 ersichtlich. Da die Cluster-Struktur der abhängigen Variable nun aufgrund der Transformation der Berufsschulnoten weitgehend verschwunden ist, kann auf Cluster-robuste Standardfehler verzichtet werden. Die Ergebnisse liefern ein ähnliches Bild, wie wir es schon von den linearen Regressionsmodellen her kennen: Während die Schulzeugnisse ein valider Prädiktor sind, ist der Koeffizient des Multicheck-Gesamtergebnisses nur im Basismodell der Berufsschulnoten des dritten Semesters signifikant. Da-

[75] Eine graphische Illustration des Modells findet sich in Grant (2007).

[76] Die Informationskriterien bevorzugen – unter Verwendung des erweiterten Basismodells – im Falle der Berufsschulnoten im ersten Semester das Ordered-Probit-, im Falle der Drittsemester-Noten das Ordered-Logit-Modell.

[77] Probit-Modelle für geordnete Kategorien unterstellen den exogenen Variablen des Modells einen konstanten Effekt auf die latente Variable *lquartillschnitt**. Die Annahme kann mithilfe eines Likelihood-Ratio-Tests überprüft werden, der in Stata implementierbar ist (Long und Freese 2003). Die Idee des Tests besteht darin, die Ordered-Probit-Koeffizienten mit denselben Koeffizienten in Probit-Modellen zu vergleichen, die je für zwei Ausprägungen der kategorialen Variable einzeln geschätzten werden. Oft wird in diesem Zusammenhang auch ein Wald-Test angewendet, der allerdings in Stata nur für das Ordered-Logit-Modell zur Verfügung steht. Aufgrund der Erkenntnisse aus den linearen Regressionsmodellen ist es allerdings nicht verwunderlich, dass der LR-Test keine Anzeichen auf Verletzung der *Parallel Regression Assumption* in den Modellen anzeigt.

TABELLE 7: ORDERED-PROBIT-REGRESSIONEN DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN DES ERSTEN UND DRITTEN SEMESTERS (BASISMODELL UND ERWEITERTES BASISMODELL)

	lquartillschnitt1		*lquartillschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweitertes Basismodell
mcgesamt	0.00911	0.00496	0.0133+	0.0102
	(0.00740)	(0.00786)	(0.00743)	(0.00785)
zschnitt12	0.644**	0.991**	0.487**	0.730**
	(0.158)	(0.183)	(0.158)	(0.179)
laltera	0.169	0.124	-0.126	-0.161
	(0.184)	(0.189)	(0.186)	(0.190)
lalterj	-0.0795	-0.0922	-0.0435	-3.88e-05
	(0.257)	(0.267)	(0.257)	(0.267)
weiblich	-0.0203	-0.0583	0.0863	0.0917
	(0.135)	(0.143)	(0.136)	(0.143)
muttersp	0.298*	0.389*	0.164	0.209
	(0.147)	(0.153)	(0.150)	(0.154)
stadt	-0.0182	-0.189	-0.0644	-0.164
	(0.165)	(0.173)	(0.166)	(0.174)
bvs	0.0972	0.129	0.0707	0.0859
	(0.148)	(0.154)	(0.148)	(0.154)
sek	0.558**	0.961**	0.410**	0.744**
	(0.143)	(0.174)	(0.143)	(0.172)
gym	1.235**	2.014**	1.564**	2.140**
	(0.451)	(0.500)	(0.480)	(0.512)
luzern		0.0484		0.190
		(0.457)		(0.448)
zuerich		0.0750		0.371
		0.0484		(0.403)
α_1	3.172**	5.119**	2.573**	3.863**
α_2	4.741**	6.787**	4.212**	5.572**
Observationen	329	329	324	324
Pseudo-R-Quadrat	0.069	0.112	0.0564	0.0883

Kantonsdummies nicht dargestellt (Modell 2 und 4)
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

mit ist die Evidenz aus den multivariaten Modellen eindeutig: Der Signalgehalt der Volksschulnoten ist in Bezug auf Berufsschulnoten deutlich grösser, als jener des Multicheck-Gesamtergebnisses.

Auch an der Signifikanz der Kontrollvariablen ändert sich kaum etwas: Weiterhin sind Schultyp und für das erste Semester die Muttersprache prognostisch aussagekräftig, währendem Alter, der Besuch einer berufsvorbereitenden Schule, die städtische Herkunft sowie die Zugehörigkeiten zu einer Migros-Genossenschaft weder Vor- noch Nachteile in der Berufsschule zu bringen scheinen. Die gezeigten Gütemasse – es handelt sich um McFadden's Pseudo-R^2 (Greene 2003; Long und Freese 2003) – deuten darauf hin, dass die erweiterten Standardmodelle gegenüber den Basismodellen zu favorisieren sind. Im Gegensatz zur ersten Evidenz aus den linearen Regressionsmodellen, nimmt die Aussagekraft der Schulnoten im Ordered-Probit-Modell eher ab – wiederum sind die Unterschiede der Punktschätzer für *zschnitt12* aber nicht signifikant. Alles in allem bleibt es folglich unmöglich, genauere Aussagen über den zeitlichen Verlauf des Signalwerts der Schulnoten zu machen. Der erwartete Befund, der Signalgehalt werde mit der Zeit abnehmen, kann nicht gestützt werden.

Wie sind die geschätzten Koeffizienten zu interpretieren? Es handelt sich um ein nichtlineares Modell in den Wahrscheinlichkeiten der abhängigen Variable (vgl. Abbildung 9), weshalb die Koeffizienten per se wenige Informationen über das Ausmass ihrer Effekte geben. In unserem Zusammenhang ist es am sinnvollsten, die Resultate anhand individuell prognostizierter Wahrscheinlichkeiten, in den drei Kategorien der abhängigen Variable zu landen, darzustellen (siehe Long und Freese 2003). Dabei sind die Ergebnisse konditional auf die anderen Regressoren. Deshalb müssen diesen zur Analyse spezifische Werte vorgegeben werden. So enthält Tabelle 8 Wahrscheinlichkeiten, zum ersten, den mittleren sowie dem vierten Quartil der Notenverteilung der jeweiligen Berufsschule zu gehören, wenn die Personen die in der ersten Spalte beschriebenen Merkmale mitbringen (blau: Modell für *lquartillschnitt1*; rot: Modell für *lquartillschnitt3)*.

So ist die Wahrscheinlichkeit nur zwei Prozent, dass ein Referenzgruppenlernender (erste Zeile) mit einem durchschnittlichen Schulzeugnis und durchschnittlichen Multicheck-Gesamtergebnis im ersten Lehrsemester zum besten Quartil seiner Berufsschule zählt. Im Gegensatz dazu gehört er mit fast 64-prozentiger Wahrscheinlichkeit zum ersten Quartil. Dieser Wert sinkt auf 49 Prozent, falls es sich um einen Lernenden mit deutscher Muttersprache handelt. Ein sonst identischer Lehrling deutscher Muttersprache – mit demselben Multicheck-Resultat und denselben Noten – ist, wenn er die Sekun-

TABELLE 8: MITHILFE DER ERWEITERTEN ORDERED-PROBIT-MODELLE NACH PERSÖNLICHEN MERKMALEN GESCHÄTZTE INDIVIDUELLE WAHRSCHEINLICHKEITEN ZUM ERSTEN, DEN MITTLEREN ODER DEM VIERTEN QUARTIL DER NOTENVERTEILUNG DER JEWEILS BESUCHTEN BERUFSSCHULE ZU GEHÖREN

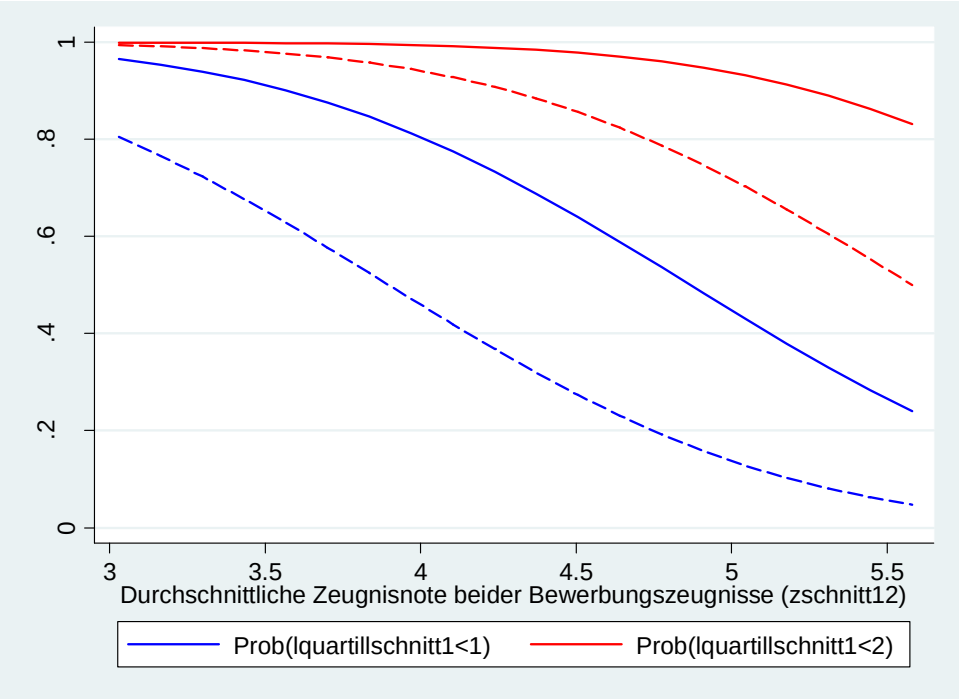
	P(*lquartillschnitt*=0|x)	P(*lquartillschnitt*=1|x)	P(*lquartillschnitt*=2|x)
Referenzgruppenlernender: Männlicher Lehrling der Migros Aare durchschnittlichen Alters und nichtdeutscher Muttersprache, der im Kanton Aargau auf dem Land in die Realschule ging und keine BVS besuchte	0.64 [0.4554, 0.8277]	0.34 [0.3267, 0.3480]	0.02 [-0.0061, 0.0484]
	0.49 [0.2397, 0.6206]	0.47 [0.4715, 0.5257]	0.05 [0.0025, 0.1401]
Referenzgruppenlernender mit deutscher *Muttersprache*	0.49 [0.2886, 0.6840]	0.46 [0.4463, 0.4870]	0.05 [0, 0.0985]
	0.40 [0.2303, 0.5775]	0.52 [0.4964, 0.5528]	0.07 [0.0072, 0.1357]
Referenzgruppenlernender mit deutscher Muttersprache, der in die *Sekundarschule* ging	0.16 [0.0620, 0.2618]	0.59 [0.5239, 0.6565]	0.25 [0.1213, 0.3745]
	0.16 [0.0624, 0.2614]	0.60 [0.5392, 0.6664]	0.24 [0.1116, 0.3590]
Weibliche Referenzgruppenlernende mit deutscher Muttersprache, die in die Sekundarschule ging	0.17 [0.0742, 0.2790]	0.59 [0.5300, 0.6572]	0.23 [0.1112, 0.3484]
	0.14 [0.0520, 0.2288]	0.60 [0.5275, 0.6628]	0.26 [0.1377, 0.3912]

Blau: Modell für *lquartillschnitt1*; Rot: Modell für *lquartillschnitt3*
In eckigen Klammern: Mit der Delta-Methode berechnete 95%-Konfidenzintervalle. Die berechneten Wahrscheinlichkeiten ergeben sich für Personen, welche im Multicheck-Gesamtergebnis sowie in den Bewerbungszeugnissen durchschnittliche Werte aufweisen (*mcgesamt*=60.17 und *zschnitt12*=4.499)

darschule besucht hat, nur mit 16-prozentiger Wahrscheinlichkeit im untersten Quartil. Der geringe Notenvorteil der Sekundarschüler von ungefähr 0.3 Notenpunkten führt wegen der relativ gestauchten Verteilung der Berufsschulnoten also bereits dazu, dass ehemalige Sekundarschüler in der Detailhandelslehre doch ein wesentlich kleineres Risiko aufweisen, zu den Schlechtesten ihrer Berufsschule zu gehören.

Für kontinuierliche Variablen lässt sich die obige Analyse über eine Skala der erklärenden Variablen ausführen. In Abbildung 9 wurde dies für den Volkschulzeugnisdurchschnitt (*zschnitt12*) getan.[78] Die berechneten Wahrscheinlichkeiten ergeben sich für die Referenzgruppe (ausgezogene Linien) und für einen Lernenden, der sich von der Referenzgruppe nur darin unterscheidet, dass er die Sekundarschule besucht hat (gestrichelte Linien), mit je einem Multicheck-Gesamtergebnis von 60 Punkten. Die Abbildung fördert deutlich zutage, dass die Wahrscheinlichkeit, sich als Lehrling im untersten Quartil der Notenverteilung der Berufsschule wiederzufinden, kontinuierlich und in relativ grossem Ausmass mit der Höhe des Zeugnisschnittes in den Bewerbungszeugnissen sinkt (blau), während die Wahrscheinlichkeit, zum vierten Quartil zu gehören, steigt. Ebenso klar tritt der Effekt des Schultyps zutage. Die Wahrscheinlichkeit, dass ein Sekundarschüler mit einem Notenschnitt von 5.5 in den Bewerbungszeugnissen zum untersten Quartil der Notenverteilung gehören wird, ist weniger als 10 Prozent, indessen er mit fast 50-prozentiger Wahrscheinlichkeit zum besten Quartil gehört.

ABBILDUNG 9: MIT DEM ORDERED-PROBIT-MODELL ÜBER DIE SCHULNOTENSKALA GESCHÄTZTE KUMULATIVE WAHRSCHEINLICHKEITEN IM ERSTEN SEMESTER DEM ERSTEN ODER DEN MITTLEREN QUARTILEN DER NOTENVERTEILUNG DER JEWEILS BESUCHTEN BERUFSSCHULE ZUZUGEHÖREN



Ausgezogene Linien: Referenzgruppenlernender. Gestrichelte Linien: Referenzgruppenlernender, der die Sekundarschule besuchte. Jeweils erweitertes Basismodell für *lquartillschnitt1*.

[78] Eine ähnliche Analyse für das Multicheck-Gesamtergebnis macht wenig Sinn, da der Koeffizient jeweils insignifikant ist.

4.3.3 ROBUSTHEIT

Die Robustheitsanalysen der linearen Modelle aus Kapitel 4.2 wurden auf das Ordered-Probit-Modell übertragen.[79] Zunächst soll auch in diesem Modell ermittelt werden, inwiefern der Signalgehalt des Multicheck-Gesamtergebnisses bei Werten über dem Stichprobenmittelwert ansteigt. Dafür wurden die Schätzungen des Modells nur für jene 155 Lernenden durchgeführt, welche im Multicheck-Gesamtergebnis mehr als 60 Punkte erreichten. Die Ergebnisse der Regressionen für diese Teilstichprobe befinden sich in Tabelle 9. Die Evidenz ist deutlicher als im Fixed-Effects-Modell: Im oberen Bereich ist das Multicheck-Gesamtergebnis im Ordered-Probit-Modell ein relativ valider Prädiktor, während er bei Gesamtergebnissen unter 60 Punkten jeweils insignifikant ist.

Wie ist dieses Resultat zu erklären? Eine naheliegende Vermutung ist, dass sich die Lehrlinge der Stichprobe, welche ein Multicheck-Ergebnis unter 60 Punkten erreichten, systematisch von Jugendlichen unterscheiden, welche im Multicheck ebenfalls weniger als 60 Punkte erreichten, aber keine Lehrstelle erhalten haben und darum nicht in der Stichprobe auftauchen.[80] Korrelieren die vorteilhaften Eigenschaften der Lehrlinge in der Stichprobe auch positiv mit den Noten in der Berufsschule, würde dies salopp gesprochen dazu führen, dass den schlechten Multicheck-Gesamtergebnissen die schlechten Berufsschulnoten systematisch wegselektioniert wurden (vgl. Multicheck 2008). Aufgrund dieser Selektionsproblematik wäre der Koeffizient des Multicheck-Gesamtergebnisses im Vergleich zum Populationsparameter gegen unten verzerrt. Dass dieser Effekt eine Rolle spielt, ist kaum von der Hand zu weisen. So dürften die Lernenden der Stichprobe gegenüber jenen, welche nicht in die Stichprobe selektioniert wurden, über grössere nichtkognitive Fähigkeiten wie Motivation, Freundlichkeit oder Offenheit verfügen – Eigenschaften, die ihnen in der Berufsschule einen Notenvorteil verschaffen.

Trotzdem ist der Einfluss des Vorselektionsproblems auf die Resultate dieser Studie zu relativieren.

TABELLE 9: ORDERED-PROBIT-REGRESSIONEN FÜR MULTICHECK-GESAMTERGEBNISSE AB 60 PUNKTEN (ERWEITERTES BASISMODELL)

	lquartillschnitt1	*lquartillschnitt3*
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.0336+	0.0400*
	(0.0182)	(0.0179)
zschnitt12	1.341**	1.010**
	(0.291)	(0.276)
Observationen	155	152
Pseudo-R-Quadrat	0.165	0.138

Restlicher Regressionsoutput nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[79] Wie im FE-Modell wurde auch das Ordered-Probit-Modell nach Multicheck-Durchführungsjahr und nach Multicheckversion einzeln geschätzt. Die Ergebnisse unterscheiden sich qualitativ nicht von denen, welche bereits im FE-Modell besprochen wurden (siehe 4.2.2).

[80] Bei Multicheck-Gesamtergebnissen über 60 Punkten ist die Gefahr dieses Selektionseffekts deutlich geringer, da jene Lernende laut Multicheck alle die beruflichen Anforderungen „übertreffen“ und damit die Bedeutung des Multichecks in der Selektionsentscheidung kleiner sein dürfte.

Einerseits sind selbst die Koeffizienten des Multicheck-Gesamtergebnisses über 60 Punkte sowohl in den Fixed-Effects- wie auch in den Ordered-Probit-Modellen – wenn sie überhaupt signifikant sind – im Vergleich zu jenen der Volksschulnoten klein. Andererseits ist nicht anzunehmen, dass die grosse Streuung der Berufsschulnoten, welche um Multicheck-Gesamtergebnisse zwischen 35 und 50 Punkten zu beobachten ist, bei Multicheck-Ergebnissen unter 35 Punkten abrupt aufhört (vgl. Abbildung 7 und Abbildung 8). Anders ausgedrückt ist es angesichts der Daten unwahrscheinlich, dass Multicheck-Gesamtergebnisse unter 40 Punkte keine guten bis sehr guten Lehrleistungen zulassen. Es ist deshalb fragwürdig, inwiefern die von Multicheck gesetzte Grenze von 40 Punkten wirklich relevant und damit auch zulässig ist.

Da Kinder mit höherem sozioökonomischen Status bei gleichen Kompetenzen systematisch höhere Volksschulnoten als Kinder aus bildungsferneren und ärmeren Verhältnissen (vgl. Abschnitt 2.2.1) erhalten[81], erscheint es angebracht, die Robustheit der bisherigen Ergebnisse auch für jene Teilstichprobe zu überprüfen, für welche der Beruf der Eltern und damit die soziale Herkunft erfasst werden konnte. Dabei ist zu beachten, dass der Einbezug der Muttersprache in den bisherigen Modellen einen Teil des Effekts des sozioökonomischen Hintergrunds – die kulturelle Herkunft – bereits berücksichtigte. Tabelle 22 in Anhang 7.F zeigt die Ergebnisse der Regressionen mit dem Ordered-Probit-Modell, in welcher die Variable der sozialen Herkunft einbezogen wurde. Dabei weist die jeweils linke Spalte die Ergebnisse im erweiterten Basismodell aus, wenn *berufe12* beigefügt wird, und die rechte stellt das Ergebnis dar, wenn das erweiterte Basismodell in der gleichen Stichprobe mit den bisher verwendeten Variablen geschätzt wird.

Die Insignifikanz des Koeffizienten dieser neu eingefügten Variable deutet darauf hin, dass die soziale Herkunft der Lernenden keinen Einfluss auf ihre Leistungen in der Berufsschule ausübt. Zudem erweisen sich die Koeffizienten von *mcgesamt* und *zschnitt12* als robust gegenüber der Einführung der Kontrollvariable. Die Punktschätzer sind quasi identisch. Einzig der Koeffizient der Muttersprache im Modell der Erstsemesterberufsschulnoten und jener des höheren Alters im Modell der Berufsschulnoten ein Jahr später verlieren aufgrund der Einführung ihre statistische Signifikanz. Desungeachtet ist der Schluss zulässig, dass die bisherige Nichtbeachtung des sozialen Hintergrunds der Lernenden die interessierenden Koeffizienten nicht verzerrte – dies umso mehr, als die Einführung von *berufe12* in weiteren Kontrollregressionen auch in den Fixed-Effects-Modellen keine wesentlichen Veränderungen der Koeffizienten und deren Signifikanz hervorruft.

4.4 Differenzierungen und Erweiterungen

Die bisherigen Resultate belegen, dass dem Schulzeugnis – Schulnoten und Schultypus – ein wesentlich grösserer Informationsgehalt zukommt, als dem Multicheck-Gesamtergebnis. Diese Erkenntnisse sind robust gegenüber der Wahl des Modells, des Schätzverfahrens oder der Kontrollvariablen. Die Erkenntnisse bedürfen aber der weiteren Einordung. So zeigt sich bezüglich dem Signalgehalt der Schulnoten, dass dieser je nach Kanton und Muttersprache, aber nicht nach Schultyp variiert (Kapitel 4.4.2) und dass Einzelnoten kontextspezifisch durchaus aussagekräftig sind (Kapitel 4.4.4).

[81] Auch in den vorliegenden Daten deutet der signifikant positive punktbiseriale Korrelationskoeffizient von 0.2 zwischen *berufe12* und *zschnitt12* auf einen Effekt in diese Richtung hin.

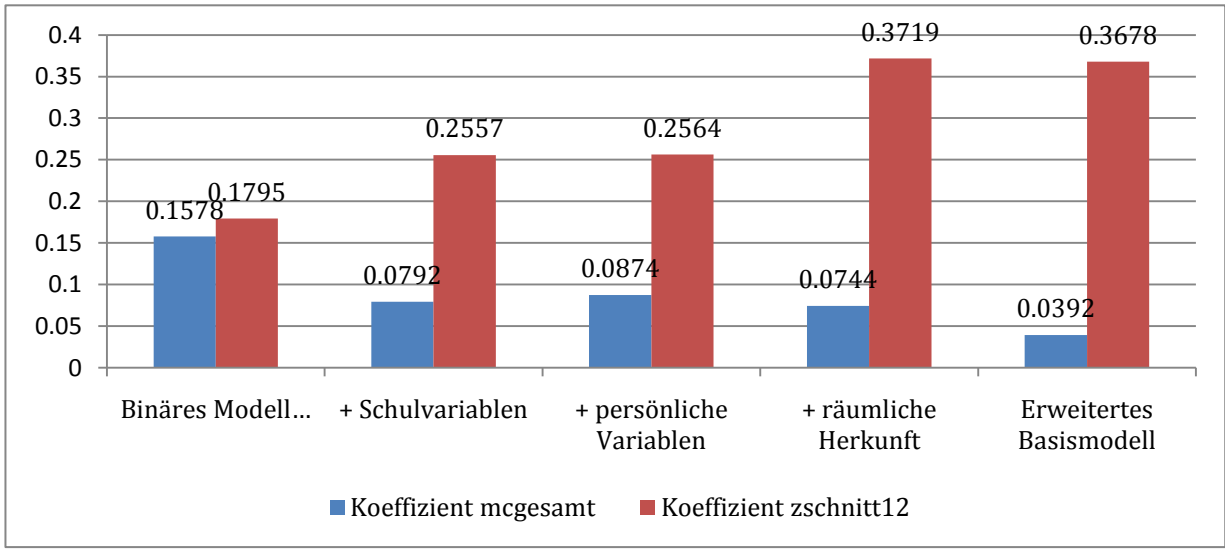
Ausserdem weisen die im Zeugnis vermerkten Berufsschulabsenzen einen positiven und teilweise eigenständigen Signalgehalt auf (Abschnitt 4.4.3). Die Resultate in Bezug auf den Multicheck sind insofern zu relativieren, als dass einzelne Multicheck-Testgebiete die entsprechende Berufsschulnote signifikant vorhersagen (Kapitel 4.4.4 und 4.4.5). Zunächst soll aber gezeigt werden, dass die Berücksichtigung kontextueller Variablen den Signalgehalt der Schulnoten vergrössert, während jener des Multichecks darunter leidet.

4.4.1 Kollinearität und Signifikanz: kontextuelle Bedingtheit des Signalgehalts

Zur Untersuchung der Kontextabhängigkeit des Signalgehalts der beiden interessierenden Variablen – *zschnitt12* und *mcgesamt* – wurde je aufbauend auf einem binären Ordered-Probit-Modell überprüft, inwiefern die Hinzunahme von Kontrollvariablen die Signifikanz und den Punktschätzer der beiden Variablen beeinflusst. Die Ergebnisse dieser Regressionen – je für *lquartillschnitt1* und *lquartillschnitt3* getrennt – sind in Anhang 7.H dargestellt (Tabelle 23 und Tabelle 24). Bei der Interpretation der Koeffizienten ist allerdings Vorsicht geboten. Sie zeigen den marginalen Effekt auf die latente Variable an, deren Metrik nicht bekannt ist. Will man die Koeffizienten der verschiedenen Modelle direkt vergleichen, müssen sie mit der geschätzten Standardabweichung der latenten Variable und jener der Koeffizienten standardisiert werden (Long und Freese 2003). Die so errechneten Koeffizienten für *mcgesamt* und *zschnitt12* im Modell für die Berufsschulnoten des ersten Semesters wurden in einem Balkendiagramm abgetragen (Abbildung 10).[82]

Die Darstellung offenbart, dass der Koeffizient des Multicheck-Gesamtergebnisses und jener der Schulnoten in einem binären Modell relativ zu ihrer jeweiligen Standardabweichung ähnlich gross sind. Der Multicheck-Koeffizient ist zudem auf dem 5%-Niveau signifikant. Sobald um den Schulty-

ABBILDUNG 10: GRÖSSE DER STANDARDISIERTEN KOEFFIZIENTEN VON MCGESAMT UND ZSCHNITT12 IM ORDERED-PROBIT-MODELL BEI HINZUNAHME VON KONTROLLVARIABLEN



Abhängige Variable: *lquartillschnitt1*

[82] Eine sehr ähnliche Darstellung liesse sich mit den standardisierten Koeffizienten für *lquartillschnitt3* erstellen. Die Koeffizienten sind folgendermassen zu interpretieren: Eine Erhöhung um eine Einheit in der erklärenden Variable k führt ceteris paribus zu einer Erhöhung um β_k Standardabweichungen in der latenten Variable *lquartillschnitt**.

pus kontrolliert wird, sinkt der Koeffizient für *mcgesamt* wie auch seine Signifikanz[83] deutlich, während er durch die persönlichen oder regionalen Kontrollvariablen nicht wesentlich beeinflusst wird. Das Ergebnis stellt einen Bezug zum ökonometrischen Modell des Multichecks her, in welchem der Schultyp das Ergebnis im Multicheck signifikant beeinflusst. Augenscheinlich verliert das Gesamtresultat einen grossen Teil seines Signalgehalts, sobald der Schultypus der Lernenden berücksichtigt wird. Oder anders herum ausgedrückt besteht ein erheblicher Teil der Aussagekraft des Multicheck-Gesamtergebnisses darin, für Schüler aus Schulzügen mit erweiterten und hohen Ansprüchen bessere Resultate auszuweisen. Werden dem Modell nun noch die Schulnoten hinzugefügt wird der Koeffizient für das Multicheck-Gesamtergebnis sehr klein und insignifikant.

Damit ist im Allgemeinen der Schluss zulässig, dass bei Berücksichtigung des Schul*zeugnisse*s der Mehrwert des Multicheck-*Gesamtergebnisses* für die Lehrlingsselektion sehr klein wird. Die Erkenntnis widerspricht auch der formulierten Hypothese, dass der Multicheck durch die Messung effektiver schulischer und kognitiver Fähigkeiten einen eigenständigen, von schulischen Leistungsausweisen und askriptiven Merkmalen grösstenteils unabhängigen Signalgehalt aufweist. Es bleibt in diesem Zusammenhang aber unklar, ob dieser Effekt auftritt, weil der Multicheck die erwähnten Grössen nicht zu messen vermag, ob sie in der Detailhandelsberufsschule nicht von Bedeutung sind, oder ob die schulischen Leistungsausweise, vor allem das Schulniveau, effektive schulische Leistungen und kognitive Fähigkeiten derart gut signalisieren, dass der Beitrag des Multichecks unbedeutend wird.

Interessanterweise verläuft die Analyse genau umgekehrt, wenn die Signifikanz und die Effektgrösse der Schulnoten betrachtet werden. Der Koeffizient der Variable *zschnitt12* wird unter Hinzunahme der Kontrollvariablen ständig grösser. Dabei sind es nicht die persönlichen Merkmale, sondern vor allem die schulische und räumliche Herkunft der Jugendlichen, welche den Signalgehalt ihrer Schulnoten erhöhen. Diese Erkenntnisse belegen den Befund der psychologischen Literatur, dass der prognostische Wert von Schulnoten stark von ihrer kontextuellen Entstehungsgeschichte abhängt (Baron-Boldt 1989). So werden unter Berücksichtigung der Muttersprache, des Schultyps oder der räumlichen Herkunft wohl vor allem die Referenzgruppenfehler und die systematischen Fehlerkomponenten in der Leistungsmessung durch Volksschulnoten korrigiert. So dürfte sich der Signalgehalt der Noten wegen der Messfehler im binären Modell grösstenteils aus seiner sozialen Komponenten speisen. Mit der Berücksichtigung weiterer Kontextvariablen werden Schulnoten anschliessend immer mehr als Leistungs- und Kompetenzmass interpretierbar.[84]

4.4.2 Schulnotensignalgehalt nach Schultyp, Kanton und Muttersprache

Einleitend wurde die Frage aufgeworfen, ob der Signalwert der Schulnoten nach Schultyp oder nach regionaler Herkunft unterschiedlich ist. So lässt sich beispielsweise die Hypothese aufstellen, dass die Realschulnoten weniger vertrauenswürdig sind als jene der Sekundarschule. Das ökonometri-

[83] Gemessen am p-Wert. Vgl. Modelle 2 und 6, Tabelle 23.

[84] Die Ergebnisse dieses Abschnitts wurden mit dem FE-Modell auf ihre Robustheit überprüft. Die Resultate unterscheiden sich kaum. Auch hier wird der Koeffizient von *mcgesamt,* je nach Spezifikation, insignifikant oder wesentlich kleiner, wenn die Variablen des Schultyps und die Kantons-Dummy-Variablen berücksichtigt werden, während der Koeffizient von *zschnitt12* gerade dann stark an Signalgehalt gewinnt.

sche Standardverfahren zum Testen solcher Fragestellungen ist das Bilden von Interaktionstermen zwischen den betreffenden Variablen. Wie Ai und Norton (2003) in einem oft zitierten Paper ausführen, ist die Inferenz und die Interpretation der Interaktionsterme bei der Schätzung eines nichtlinearen Wahrscheinlichkeitsmodells mit Schwierigkeiten verbunden und nicht einfach von der normalen, linearen Regression übertragbar.[85] Deshalb wird in diesem Abschnitt auf das Fixed-Effects-Modell zurückgegriffen. Dies hat keine wesentlichen Nachteile, da sich das Modell als gut spezifiziert erwiesen hat.

Zunächst werden die Schulnoten mit dem Schultyp (*sek*) interagiert, wodurch sich der prädiktive Gehalt von Sekundarschulnoten aus zwei Koeffizienten, jenem für *zschnitt12* und jenem für *zschnitt12sek*, zusammensetzt.[86] Tabelle 25, Anhang 7.I enthält die Ergebnisse dieser FE-Regressionen. Man erkennt an der Insignifikanz des Interaktionsterms, dass sich die Validität der Schulnoten nicht nach Schultyp unterscheidet. Die Punktschätzer legen allerdings nahe, dass Sekundarschulnoten möglicherweise etwas aussagekräftiger als Realschulnoten sind. Nichtsdestotrotz legitimiert diese Feststellung das in dieser Studie verwendete Vorgehen, den Signalgehalt der Schulnoten innerhalb einer einzelnen Variable zu analysieren und nicht nach Schulsystemen zu trennen.

Analog zu den Schultypen können die Noten auch mit den Kantons-Dummy-Variablen interagiert werden. Im Gegensatz zur vorangehenden Schätzung wurden alle Schulnoten interagiert, so dass kein separater Koeffizient für *zschnitt12* übrig bleibt. Dabei wurden die Schulnoten aus jenen Kantonen, aus welchen weniger als 50 Lehrlinge in der Stichprobe stammten, im residualen Interaktions-

TABELLE 10: SCHULNOTENVALIDITÄT NACH KANTONALER HERKUNFT DER LERNENDEN

	lschnitt1	*lschnitt3*
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
zschnitt12be	0.377**	0.553*
	(0.0676)	(0.207)
zschnitt12ag	0.426**	0.418**
	(0.0672)	(0.0792)
zschnitt12lu	0.330**	0.357**
	(0.0392)	(0.0848)
zschnitt12divkan	0.328**	0.306*
	(0.0514)	(0.130)
zschnitt12zh	0.165**	0.228*
	(0.0530)	(0.0953)
Observationen	329	324
R-Quadrat	0.309	0.268
Cluster	23	23

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[85] Eine Möglichkeit zur Lösung des Problems, die von Buis (2010) kürzlich aufgezeigt wurde, ist die Betrachtung der Odds Ratio, für welche die Inferenz wie auch die Interpretation der Parameter direkt aus dem Regressionsoutput abgelesen werden kann – mit dem Nachteil einer schwierigeren Interpretierbarkeit. Allerdings ist die Berechnung der Odds Ratio für ein Ordered-Probit-Modell in der verwendeten Stata-Version nicht implementiert.

[86] Aufgrund der geringen Fallzahl der Lernenden, welche eine Schule mit hohen Ansprüchen besucht haben, wird auf eine Berücksichtigung in der Schätzung verzichtet.

term *zschnitt12divkan* zusammengefasst.

Die Punktschätzer dieser Interaktionsterme unterscheiden sich beträchtlich (vgl. Tabelle 10). Bei Lernenden aus dem Kanton Bern, die einen um eine Note höheren Notendurchschnitt in den Bewerbungszeugnissen aufweisen, ist ceteris paribus mit einem um 0.553 Notenpunkten besseren Schnitt im dritten Berufsschulzeugnis zu rechnen. Die gleiche Differenz führt in einem Bewerbungszeugnis aus dem Kanton Zürich nur zu einem geschätzten Unterschied von 0.228 Notenpunkten in der Berufsschule. Allerdings sind aufgrund der relativ grossen Standardabweichungen nur wenige Unterschiede in den Koeffizienten statistisch signifikant. So unterscheiden sich auf dem 10%-Signifikanzniveau im Modell für die Berufsschulnoten des ersten Semesters nur die Koeffizienten zwischen *zschnitt12ag* beziehungsweise *zschnitt12be* und *zschnitt12zh* und im Modell für die Berufsschulnoten des dritten Semesters nur jene zwischen *zschnitt12ag* und *zschnitt12zh*.[87]

Ein interessantes Ergebnis liefert die Interaktion der Schulnoten mit der Muttersprache der Lernenden: Die Aussagekraft der Schulnoten von Lernenden deutscher Muttersprache ist fast um das Doppelte grösser, als jene der Schulnoten der Lernenden nichtdeutscher Muttersprache (siehe Tabelle 25).[88] Wie ist dieses Resultat zu erklären? Möglicherweise sind die Volksschulnoten Jugendlicher nichtdeutscher Muttersprache deshalb besonders fehlerbehaftet, weil die Noten jener Jugendlichen im Vergleich zur Referenzgruppe noch weniger von ihrem tatsächlichen Leistungsvermögen abhängen (bspw. Kronig 2007). Insofern diese Beurteilungsfehler in der Berufsschule nicht fortbestehen, wäre der Signalgehalt der Schulnoten für jene Jugendlichen kleiner. Andererseits könnte der Effekt auch aufgrund der unterschiedlichen Lehrpläne von Volks- und Berufsschule zustande kommen – zum Beispiel, wenn sich sprachliche Defizite in den neuen Fächern in der Berufsschule in kleinerem

TABELLE 11: SCHULNOTENVALIDITÄT NACH MUTTERSPRACHE DER LERNENDEN

	lschnitt1	*lschnitt3*
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
zschnitt12	0.188*	0.216*
	(0.0697)	(0.0860)
zschnitt12muttersp	0.154*	0.186*
	(0.0709)	(0.0792)
Observationen	329	324
R-Quadrat	0.303	0.264
Cluster	23	23

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[87] Zur Überprüfung dieser Resultate wurden auch Interaktionsterme zwischen Schulnoten und den Regionen-Dummy-Variablen geschätzt. Dieses Vorgehen wurde gewählt, um die Hypothese zu überprüfen, dass sich die drei Migros-Genossenschaften darin unterscheiden, schulisches Vorwissen in Berufsschulleistungen zu überführen. Die Punktschätzer der Interaktionsterme sind aber in allen getesteten Modellen sehr ähnlich. Somit scheinen die festgestellten, kantonalen Unterschiede im Signalwert der Schulnoten durchaus teilweise auf die Differenzen in den kantonalen Schulsystemen zurückzuführen zu sein.

[88] Regressionen mit Interaktionen zwischen Volksschulnoten und der Variable *berufe12* legen nahe, dass dieser Effekt nicht auf einen indirekten Effekt der sozialen Herkunft zurückzuführen ist. Der Effekt stammt auch nicht von den kantonalen Differenzen im Signalwert der Schulnoten: Die Lernenden nichtdeutscher Muttersprache verteilen sich über alle grösseren Kantone einigermassen gleichmässig.

Umfang kumulativ auf alle Schulnoten auswirken, als es in den Fächern der Volksschule der Fall ist (Moser 2004c).[89]

4.4.3 *Signalgehalt im Schulzeugnis vermerkter Absenzen*

Die Informationen, welche die Schulzeugnisse für die Betriebe bereit stellen, beschränken sich oftmals nicht bloss auf die Schulnoten und den Schultypus. In vielen Kantonen ist es üblich, in Schulzeugnissen Betragen und soziales Verhalten der Schüler zu bewerten. Zudem weisen viele Zeugnisse auch die entschuldigten und unentschuldigten Unterrichtsabsenzen der Schüler aus. Diese Informationen werden von den Betrieben erwiesenermassen in der Lehrlingsselektion berücksichtigt (Hupka et al. 2006).[90] Für die Betriebe in Schmid und Stornis (2004) Umfrage sind unentschuldigte Schulabsenzen gar die wichtigste Informationsquelle unter den schulischen Leistungsnachweisen. Weisen diese Informationen aber tatsächlich einen für die Betriebe verwertbaren, positiven Signalgehalt auf?

Leider stand einer Erfassung der Betragens- und Verhaltensbewertungen in den Zeugnissen im Wege, dass sich die Kantone stark darin unterscheiden, welche Charakteristika sie bewerten, in welcher Detailliertheit sie diese beurteilen und in welcher Form sie diese im Zeugnis vermerken. Da die Vergleichbarkeit der Ergebnisse nicht hätte gewährleistet werden können, wurde auf eine Erfassung verzichtet. Erhoben wurden hingegen die schulischen Absenzen. Allerdings war die Erfassung der Werte ebenfalls nur unter gewissen Einschränkungen möglich. So fehlen in den Zeugnissen gewisser Kantone die Angaben vollständig (Aargau, Solothurn, Zürich, St. Gallen und Glarus) oder es fehlten

TABELLE 12: ORDERED-PROBIT-REGRESSIONEN DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN DES ERSTEN SEMESTERS UNTER VERWENDUNG DER IN DEN SCHULZEUGNISSEN AUSGEWIESENEN ENTSCHULDIGTEN ABSENZEN (ZENTSCH01)

	lquartillschnitt1			
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell inkl. *zentsch01*	Erweitertes Basismodell inkl. *zentsch01*
mcgesamt	0.0115	0.00749	0.0122	0.00820
	(0.00999)	(0.0108)	(0.0101)	(0.0109)
zschnitt12	0.872**	1.010**	0.806**	0.936**
	(0.279)	(0.300)	(0.280)	(0.301)
zentsch01			-0.379+	-0.424*
			(0.201)	(0.209)
Observationen	158	158	158	158
Pseudo-R-Quadrat	0.0924	0.117	0.104	0.130

Restlicher Regressionsoutput nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[89] Interaktionsterme zwischen Schulnoten und der Stadt-Land-Variable sowie zwischen Schulnoten und Geschlecht sind insignifikant. Insofern scheint die Validität der Schulnoten nicht davon abzuhängen, ob das Zeugnis in einer städtischen oder einer ländlichen Volksschule ausgestellt wurde oder von einer Frau oder einem Mann stammt.

[90] Ältere Studien aus den USA vermögen allerdings keinen Zusammenhang zwischen Signalen guter Verhaltensweisen in High Schools und dem frühen Arbeitsmarkterfolg aufzuzeigen (Bishop 1994).

jene Seiten des Zeugnisses, auf welcher die Absenzen vermerkt sind (Bern).[91] Damit konnten die Informationen zu den Schulabsenzen insgesamt nur für 158 Lernende erfasst werden. Schliesslich wiesen nur vier erfasste Schulzeugnisse unentschuldigte Absenzen aus.[92]

Aufgrund der kleinen Fallzahl an unentschuldigten Absenzen soll das Augenmerk auf die entschuldigten Abwesenheiten gerichtet werden. Zur Operationalisierung wurden die entschuldigten Absenzen der beiden erfassten Schulzeugnisse addiert. Anschliessend wurde aufgrund der Zensiertheit der Variable – ein beachtlicher Teil der Lehrlinge verzeichnete in der Schule gar keine Absenzen –, und um nicht Linearität der Effekte zu unterstellen, eine binäre Variable *zentsch01* gebildet. Die Variable nimmt den Wert eins an, falls der Lernende überdurchschnittlich viele halbtägige Absenzen – das heisst mehr als 7.467 – in der Schule aufweist. Diese so codierte Variable ist signifikant negativ mit den Schulnoten *zschnitt12* korreliert ($\rho = -0.176$), aber unkorreliert mit dem Multicheck-Gesamtergebnis. Zwar sind die Koeffizienten der Variable *zentsch01* in den Modellen der Berufsschulnoten des dritten Semesters nicht signifikant[93], doch erweist sich die Variable im Ordered-Probit-Modell als valider Prädiktor der Berufsschulnoten im ersten Semester (vgl. Tabelle 12). Die Koeffizienten des Multicheck-Gesamtergebnisses und der Schulnoten bleiben dabei fast unverändert. Dabei lässt sich die Grösse des Koeffizienten von *zentsch01* am besten anhand eines diskreten Wechsels der Variable illustrieren: Hatte der Lernende in der Schule überdurchschnittlich viele entschuldigte Absenzen, steigt die Wahrscheinlichkeit, dass er in der Berufsschule im ersten Semester zum untersten Quartil gehört, um 16.1 Prozent. Gleichzeitig sinkt die Wahrscheinlichkeit, zum besten Quartil zu gehören, um 10.65 Prozent. Da durch das Hinzufügen der entschuldigten Absenzen zudem McFadden's Pseudo-R^2 ansteigt, scheint sich der Signalwert eines Schulzeugnisses durch das Ausweisen der Absenzen zu erhöhen.

4.4.4 Signalwert einzelner Schulnoten und einzelner Multicheck-Testgebiete

Bislang wurde ausschliesslich der Signalwert der Zeugnisdurchschnittsnoten betrachtet. Es ist aber bekannt, dass der Informationsgehalt der Schulnoten durch Mittelwertbildung grösser wird (Baron-Boldt 1989). Grant (2007) spricht von einem „substanziellen" Effekt. Die Mittelwertbildung verringert die Verzerrungen, die zum Beispiel durch einzelne schlechte Prüfungsergebnisse oder einzelne Lehrer verursacht werden. Diese Erkenntnisse werfen die Frage auf, ob einzelne Fächernoten für sich genommen prognostisch unbrauchbar sind. Spezifischer interessiert, inwiefern aus zwei Be-

[91] Zudem mussten einige Werte von den angegebenen Lektionen in die übliche Form der Halbtage überführt werden. Die angegebene Anzahl Lektionen wurde dabei einfachheitshalber durch vier geteilt.

[92] Das dürfte mehrere Gründe haben: So kommt den unentschuldigten Absenzen in der Lehrlingsauswahl der Migros-Genossenschaften eine gewisse Bedeutung zu, weshalb Personen mit unentschuldigten Absenzen nur in seltenen Fällen in der Stichprobe auftauchen. Berücksichtigt man, dass die penalisierende Wirkung unentschuldigter Absenzen auch den Lehrern bewusst sein dürfte, ist es gut vorstellbar, dass die Lehrer diese aus Rücksicht auf die späteren Berufschancen ihrer Schüler nicht eintragen.

[93] Da die Hinzunahme von *zentsch01* auch die Koeffizienten von *mcgesamt* und *zschnitt12* in jenen Regressionen nicht beeinflusst, wird auf eine Darstellung der Ergebnisse dieser Regression verzichtet.

werbungszeugnissen zusammengenommene, einzelne Fächernoten den Notendurchschnitt, und, in Abschnitt 4.4.5, einzelne Berufsschulnoten zu prognostizieren vermögen.[94]

Dazu wurden wegen der Multikollinearität der Schulnoten[95] (Imdorf 2005; Mühlemann et al. 2007) die Noten je einzeln in die erweiterten Ordered-Probit-Basismodelle der ersten und dritten Lehrsemesterzeugnisse eingefügt und die Koeffizienten anschliessend in Tabelle 13 abgetragen.[96] Die Erkenntnisse von Grant (2007) bestätigen sich: Die geschätzten Punktschätzer der einzelnen Noten liegen deutlich unter den 0.991 (*lquartillschnitt1*) beziehungsweise 0.73 (*lquartillschnitt3*), welche in Tabelle 7 für *zschnitt12* gemessen wurden. Trotzdem ist jede Fächernote mit Ausnahme der Deutschnote für sich genommen ein signifikanter Prädiktor beider Lehrzeugnisse. In diesen Regressionsergebnissen findet sich zudem zum ersten Mal eine eindeutigere Tendenz, dass die Validität der Schulnoten im Verlauf der Lehre abnimmt, da alle Punktschätzer für *lquartillschnitt3* unter jenen liegen, die für das erste Semester geschätzt werden. Ausserdem bestätigt sich ein bekanntes Ergebnis (vgl. Baron-Boldt 1989): Die Mathematiknote ist auch bei Lehrplänen, in welchen Mathematik im engeren Sinne nicht unterrichtet wird, ein guter Prädiktor der Leistungen.

Dasselbe Experiment wurde auch für die einzelnen Multicheck-Testgebiete durchgeführt, auch hier aufgrund der Kollinearität der einzelnen Testteile je einzeln (vgl. Tabelle 14)[97]: Die Desaggregation des Multicheck-Gesamtergebnisses in die einzelnen Testgebiete offenbart, dass das Englisch- und der Merkfähigkeitsergebnis zumindest auf dem 10%-Signifikanzniveau statistisch signifikant mit den Berufsschulnotendurchschnitt zusammenhängen. Die beiden Testgebiete sind damit überraschenderweise die besseren Prädiktoren als das Gesamtergebnis. Mit anderen Worten: Es scheint mehr Sinn zu ergeben, sich in der Selektionsentscheidung auf das Englisch- und Merkfähigkeitsergebnis, als sich auf das Gesamtergebnis zu verlassen. Alle anderen Koeffizienten indes – insbesondere auch

TABELLE 13: KOEFFIZIENTEN EINZELNER ÜBER ZWEI SEMESTER GEMITTELTER FÄCHERNOTEN IM ORDERED-PROBIT-MODELL DER DURCHSCHNITTSNOTEN IM ERSTEN UND DRITTEN LEHRSEMESTER (ERWEITERTES BASISMODELL)

	Deutsch (*zd12*)	**Französisch (*zf12*)**	**Englisch (*ze12*)**	**Mathematik (*zmath12*)**	**Geschichte/ Geographie (*zgsgeo12*)**	**Naturwissenschaften (*znat12*)**
lquartillschnitt1 (N=282)	0.429* (0.176)	0.397** (0.125)	0.449** (0.130)	0.430** (0.126)	0.535** (0.145)	0.561** (0.134)
lquartillschnitt3 (N=277)	0.123 (0.175	0.234+ (0.123)	0.390** (0.130)	0.316* (0.126)	0.430** (0.145)	0.468** (0.134)

Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[94] In einigen Fällen war das gleiche Fach nicht in beiden erfassten Bewerbungszeugnissen benotet worden. In solchen Fällen wurde die zur Verfügung stehende Einzelnote berücksichtigt.

[95] Die Korrelationskoeffizienten der einzelnen Fächernoten der Volksschulzeugnisse liegen zwischen 0.38 und 0.71.

[96] Die Stichprobe umfasst jeweils dieselben 282 bzw. 277 Observationen. Der Verlust der Beobachtungen entsteht, weil nicht für alle Lehrlinge sämtliche aufgeführten Noten vorhanden waren.

[97] Hier konnte die Stichprobe allerdings nicht fixiert werden, da sich die Testteilnehmer zwischen Englisch und Französisch entscheiden können und oft nur eine Fremdsprache gewählt wird. Würde das Sample auf jene Personen restringiert, die beide Teile gemacht haben, hätten nur 177 Observationen berücksichtigt werden können. Es wurden deshalb auf Kosten der Vergleichbarkeit jeweils alle zur Verfügung stehenden Multicheck-Resultate verwendet.

die Koeffizienten derjenigen Testteile, welche ähnlich einem IQ-Test das Potenzial der Prüflinge messen beziehungsweise zu messen versuchen – sind insignifikant.

Im Übrigen haben auch die in der Multicheck-Auswertung (vgl. Anhang 7.A) ausgewiesenen Indikatoren des „Schulwissens" und des „Potenzials" keine Aussagekraft zur Prognose der Lehrdurchschnittsnoten. Die Ergebnisse bezüglich des Potenzial-Teils des Tests sind insofern überraschend, als psychologische Studien immer wieder einen signifikanten Zusammenhang zwischen Potenzialtests und einer ganzen Reihe von Erfolgsvariablen belegen konnten (siehe Kapitel 2.3). Es muss im Rahmen dieser Arbeit allerdings unbeantwortet bleiben, woher der nicht gefundene Zusammenhang zwischen dem Multicheck-Potenzialteil und den Berufsschulnoten rührt. So könnte es sein, dass der Multicheck das tatsächliche Potenzial der Testteilnehmer nicht abzubilden vermag – oder aber dass er es zwar misst, es aber für die Berufsschulnoten keine bedeutende Rolle spielt. Evidenz für den zweiten Effekt – dass nämlich weniger die kognitiven Fähigkeiten, sondern Motivation und Lernbereitschaft gute Berufsschulnoten determinieren – liefert die Studie von Stamm et al. (2009).

4.4.5 Modelle einzelner Berufsschulnoten

Die Ergebnisse aus dem vorangehenden Abschnitt belegen: Auch Einzelnoten können für die Prognose der durchschnittlichen Berufsschulleistung herangezogen werden. Allerdings gewichten die Lehrlingsausbildner, je nach Relevanz und Aussagekraft des Faches für die Berufsschule, gewisse Zeugnisnoten stärker als andere. So erachten die Lehrlingsausbildner der Migros gute Noten in Deutsch und Mathematik als wichtiger als gute Noten in den Nebenfächern. Wie steht es also um die Voraussagevalidität der Schulnoten und der Multicheck-Testgebiete, wenn es darum geht, einzelne Berufsschulnoten zu prognostizieren?

Um diese Frage zu untersuchen, wurden Modelle geschätzt, in denen als Linkhandvariable jeweils der Mittelwert der Fächernoten aus den beiden Lehrzeugnissen verwendet wird. In den Modellen wurden die exogenen Kontrollvariablen des erweiterten Basismodells um gewisse erklärende Schulnoten und Multicheck-Testteile ergänzt.[98] Diese zusätzlichen Variablen wurden anhand ihrer Relevanz und in einem zweiten Schritt aufgrund der Parametersignifikanz und der Informationskriterien

TABELLE 14: KOEFFIZIENTEN EINZELNER MULTICHECK-TESTGEBIETE IM ORDERED-PROBIT-MODELL DER DURCHSCHNITTSNOTEN IM ERSTEN UND DRITTEN LEHRSEMESTER (ERWEITERTES BASISMODELL)

	Deutsch (*mcd*)	**Französisch (*mcf*)**	**Englisch (*mce*)**	**Rechnen (*mcrechnen*)**	**Logik (*mclogik*)**	**Konzentration (*mckonz*)**	**Merkfähigkeit (*mcmerk*)**
lquartillschnitt1	-0.0014 (0.0045) N=329	0.0016 (0.0048) N=209	0.0067+ (0.0039) N=295	-0.0006 (0.0038) N=329	-0.0027 (0.004) N=329	0.0036 (0.0041) N=329	0.0075* (0.0035) N=329
lquartillschnitt3	0.0006 (0.0045) N=324	0.0033 (0.0048) N=203	0.0094* (0.0039) N=290	0.0027 (0.0038) N=324	-0.0015 (0.0040) N=324	0.00337 (0.0041) N=324	0.0069* (0.0069) N=324

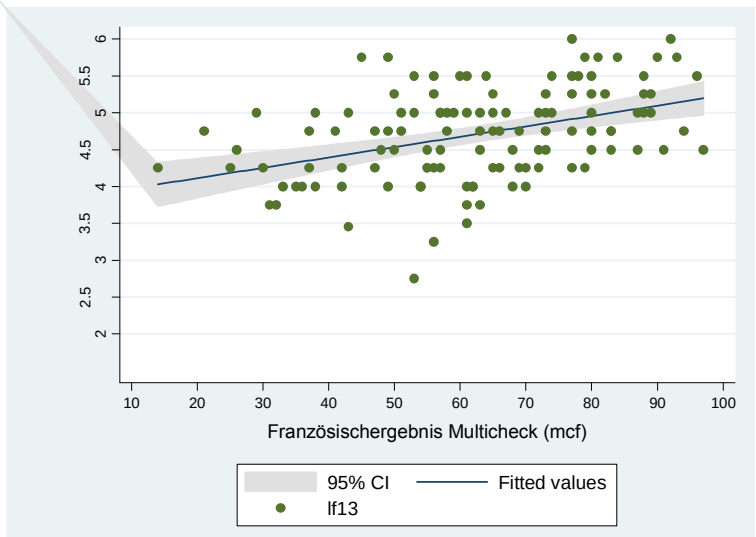
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[98] Falls, zum Beispiel aufgrund eines Lehrabbruchs, nur die Note eines Lehrzeugnisses zur Verfügung stand, wurde diese einzeln verwendet. Ebenfalls auf einer Einzelnote basiert die Note im Fach „Branchenkunde", da dieses Fach in der Berufsschule nur während eines Semesters unterrichtet wird.

ausgewählt. Die resultierenden Modelle werden jeweils den Ergebnissen einer Regression gegenübergestellt, in denen neben den Kontrollvariablen nur der Zeugnisschnitt der Bewerbungszeugnisse und das Multicheck-Gesamtergebnis als exogene Variable eingesetzt werden (*zschnitt12* und *mcgesamt*). Der Einfachheit und Interpretierbarkeit halber wurden Fixed-Effects-Modelle geschätzt.[99] Die Ergebnisse der Regressionen sind in den Tabellen im Anhang 7.J abgedruckt. Es wird in diesem Abschnitt darauf verzichtet, die Resultate einzeln und detailliert zu diskutieren. Einige Regularitäten und Erkenntnisse aus den Regressionen sind aber bemerkenswert und erweitern die bisherigen Resultate um drei interessante Aspekte.

Erstens haben jeweils jene Schulnoten den höchsten Signalgehalt, für welche es aufgrund ihres inhaltlichen Bezuges zu erwarten war: Die Deutschnote der Berufsschule wird am besten anhand der Deutschnote im Volksschulzeugnis vorausgesagt und die Note im Fach Gesellschaft durch jene in Geschichte und Geographie. Zweitens sind die aus Einzelnoten und einem Multicheck-Testgebiet gebildeten Modelle den Modellen mit *zschnitt12* und *mcgesamt* prädiktiv überlegen. Sie klären mehr Varianz in der abhängigen Variable auf, und auch die Informationskriterien sind jeweils wesentlich kleiner. Mit anderen Worten: Gezielt eingesetzt sind einzelne Noten und einzelne Multicheck-Testgebiete dem Zeugnisschnitt und dem Multicheck-Gesamtergebnis als Prädiktoren einzelner Berufsschulnoten vorzuziehen. Drittens zeigt sich, dass einzelne Testgebiete des Multichecks durchaus prädiktiv valide sind. So sind die Koeffizienten der Sprachenteile des Multichecks jeweils signifikante Prädiktoren der jeweils entsprechenden Sprachnote in der Berufsschule. Weiter hängt das Ergebnis im Bereich Merkfähigkeit signifikant mit der Note im Fach Detailhandelskenntnisse zusammen, das Ergebnis im Gebiet Rechnen ist zur Vorhersage zweier Berufsschulnoten (Wirtschaft und Branchenkunde) einsetzbar, und das Englischergebnis ist ein geeigneter Prädiktor der Berufsschulnoten in den Fächern Wirtschaft und Gesellschaft. Dabei sind die Punktschätzer der Koeffizienten zuweilen relativ gross. Illustrieren lässt sich dieser beachtliche Zusammenhang beispielsweise anhand der

ABBILDUNG 11: TATSÄCHLICHE UND GEFITTETE WERTE EINER BINÄREN REGRESSION DER DURCHSCHNITTLICHEN FRANZÖSISCHNOTEN AUF DAS FRANZÖSISCHERGEBNIS IM MULTICHECK



[99] Schliesslich wurden die Modelle nicht zuletzt für die Berufsbildner der Migros geschätzt.

gemeinsamen Verteilung der Berufsschulnote in Französisch (*lf13)* und des Französischergebnisses im Multicheck (*mcf)* (Abbildung 11).[100]

Insgesamt erhärtet sich der Eindruck, dass einzelne Multicheck-Testteile im Signalwert dem Gesamtergebnis – zumindest für die hier betrachtete Stichprobe und in Bezug auf Berufsschulnoten – überlegen sind. Das ist bei einem Test, in dem für die sechs oder wahlweise sieben Testgebiete im Schnitt zwanzig Minuten zur Verfügung stehen, doch überraschend. Dabei dürften zwei Effekte eine Rolle spielen: Einerseits sind die Verteilungen der Resultate in den einzelnen Multicheck-Testteilen weniger gegen unten „zensiert", als es das Multicheck-Gesamtergebnis ist. Das heisst, der Multicheck taugt nun als Prädiktor, weil auch sehr tiefe Testresultate – Ergebnisse unter 35 Punkten – in der Stichprobe auftauchen (vgl. Abbildung 11).

Andererseits ist die grössere prädiktive Validität der einzelnen Testgebiete wohl aber auch auf Mängel in der konzeptionellen und inhaltlichen Ausarbeitung des Tests zurückzuführen. Es scheint, als ob der Test den spezifischen Anforderungen der Detailhandelslehre in Bezug auf die Berufsschule zu wenig Rechnung trägt. So zeigen die Regressionen, dass jeweils nur eines von sechs Testgebieten als signifikanter Prädiktor für die Vorhersage der Berufsschulfächer Detailhandel, Branchenkunde und Gesellschaft eingesetzt werden kann. Dies dürfte einer der Hauptgründe sein, wieso das Multicheck-Gesamtergebnis (*mcgesamt*) einen relativ geringen Zusammenhang mit den durchschnittlichen Berufsschulnoten aufweist – zumal die Detailhandelskenntnisse doppelt gewertet werden.

Zudem zeigte sich in den Regressionen zur Modellauswahl, dass der Multicheck zwei Testgebiete enthält (Logik und Konzentration), welche keine einzige Berufsschulnote, geschweige denn den Notendurchschnitt, signifikant vorherzusagen vermögen. Dies unterminiert das Gesamtergebnis als Prädiktor der durchschnittlichen Berufsschulnoten weiter. Stimmen diese Ergebnisse, lässt sich die Berücksichtigung der Testteile im Test nur dann rechtfertigen, wenn sie einen positiven Signalwert in Bezug auf praktische Ausbildungsleistungen aufweisen.

4.5 Andere Kriterien des Lehrerfolgs

Eines verbleibt aber trotz des Fakts, dass die Testgebiete des Detailhandelsmultichecks Informationen enthalten, die zur Prognose einzelner Berufsschulfächer eingesetzt werden können, offensichtlich: Schulnoten haben zur Vorhersage von Berufsschulnoten zumindest für die Migros die klar grössere Aussagekraft. Aufgrund der theoretischen Überlegungen aus Kapitel 2 war dies in dieser Deutlichkeit nicht zu erwarten. Inwiefern sich die geäusserten Erwartungen bezüglich der beiden anderen betrachteten, abhängigen Variablen – die Wahrscheinlichkeit unentschuldigter Berufsschulabsenzen und von Lehrvertragsauflösungen – erfüllen, wird im kommenden Abschnitt erläutert.

4.5.1 Unentschuldigte Berufsschulabsenzen

Unentschuldigte schulische Absenzen sind für die Lehrbetriebe ein grosses Ärgernis. Sie deuten auf Unzufriedenheit, emotionale Instabilität, fehlendes Pflichtbewusstsein und Motivationsprobleme der

[100] Der Entscheid, welche Fremdsprache man im Test prüfen lässt, liegt beim Testteilnehmer und ist damit endogen. Dies muss bei der Interpretation berücksichtigt werden, denn es ist anzunehmen, dass der Teilnehmer jeweils diejenige Fremdsprache wählt, in welcher er sich relativ stärker einschätzt.

Lernenden hin (Lounsbury et al. 2004) und sind somit möglicherweise Ausdruck einer negativen Ausbildungssituation. In den Augen der Berufsbildner der Migros gilt es deshalb, unentschuldigte Berufsschulabsenzen möglichst zu vermeiden. Treten sie auf, wird mit den betroffenen Lehrlingen meist das Gespräch gesucht und allenfalls eine Verwarnung ausgesprochen. Schon nur deshalb bedeuten sie finanziellen Aufwand für den Betrieb. Da das Auftreten unentschuldigter Berufsschulabsenzen im Rahmen der Arbeit erfasst werden konnte, soll im Folgenden untersucht werden, inwiefern die Schulnoten und das Multicheck-Gesamtergebnis zur Prognose der Abwesenheiten eingesetzt werden können. Die abhängige Variable ist dabei dichotom. Zwar ist in den meisten Berufsschulzeugnissen die genaue Anzahl an unentschuldigten Absenzen angegeben und damit der Einsatz eines Tobit-Modelles angezeigt, doch weisen die Berufsschulen in der Region Zürich die Absenzen in ihren Zeugnissen nicht aus. Da eine Nichtberücksichtigung jener Beobachtungen eine zu grosse Einschränkung der Stichprobengrösse bedeutet hätte, wurden die zuständigen Lehrlingsbetreuer gebeten, anzugeben, welche Lehrlinge unentschuldigte schulische Absenzen verzeichneten. Darauf aufbauend wurde für die gesamte Stichprobe eine binäre Variable *lunentsch01* gebildet, welche den Wert eins

TABELLE 15: LOGIT- UND PROBIT-MODELL DER WAHRSCHEINLICHKEIT, UNENTSCHULDIGTE ABSENZEN WÄHREND DEN ERSTEN DREI LEHRSEMESTERN ZU VERSCHULDEN

	lunentsch01	
	(1)	(2)
VARIABLEN	Logit	Probit
mcgesamt	0.0114	0.00734
	(0.0218)	(0.0124)
	[0.0197]	(0.0114)
zschnitt12	-1.041*	-0.613*
	(0.460)	(0.261)
	[0.279]	[0.154]
laltera	0.304	0.179
	(0.478)	(0.268)
lalterj	0.218	0.116
	(0.733)	(0.412)
weiblich	-0.520	-0.277
	(0.370)	(0.206)
muttersp	0.0229	-0.0123
	(0.409)	(0.228)
stadt	0.545*	0.299*
	(0.412)	(0.236)
bvs	-0.145	-0.0875
	(0.412)	(0.233)
sek	-0.995*	-0.544*
	(0.477)	(0.266)
gym	-0.160	-0.0444
	(1.102)	(0.654)
Konstante	3.877*	2.200*
Observationen	290	290
Pseudo-R-Quadrat	0.166	0.166

Kantons- und Regionendummies nicht dargestellt
Standardfehler in runden Klammern,
Cluster-robuste Standardfehler in eckigen Klammern
** p<0.01, * p<0.05, + p<0.1

annimmt, wenn ein Lehrling im Verlauf der ersten drei Semester einmal oder mehrmals unentschuldigt von der Berufsschule fernblieb.

Die Standardmodelle zur multivariaten Analyse binärer abhängiger Variablen sind das Logit- und das Probit-Modell. Die Modelle haben als Spezialfall der Ordered-Logit- und Ordered-Probit-Modelle, welche bereits besprochen wurden, ebenfalls eine Interpretation mithilfe einer latenten Variable (Cameron und Trivedi 2009). Das hier geschätzte Modell ist deshalb bis auf die binäre statt kategorielle abhängige Variable identisch mit dem Modell in (4.2). Auch im hiesigen Kontext gelangen die Kontrollvariablen des erweiterten Basismodells zur Anwendung.[101] Logit- und Probit-Modell unterscheiden sich wiederum in der Verteilung, welche dem Fehlerterm der Regression unterstellt wird. Wie Greene (2003) aufzeigt, sind die standardnormale und die logistische Verteilungen vor allem an den Enden verschieden. Da von 290 Lernenden nur 48 unentschuldigte Absenzen verschuldeten (16.55%), liegt hier durchaus eine solche Ecklösung vor. Damit sind die von Logit und Probit geschätzten Wahrscheinlichkeiten möglicherweise ziemlich unterschiedlich und es wird darauf verzichtet, das Modell mit nur einer Verteilung zu schätzen. Stattdessen werden die Resultate identischer Logit- und Probit-Modelle gegenübergestellt.[102]

Die Resultate der Regressionen sind in Tabelle 15 dargestellt. Logit- und Probit-Modell unterscheiden sich dabei kaum.[103] Entgegen den formulierten Erwartungen, ist der Koeffizient des Multicheck-Gesamtergebnisses nicht signifikant. Die Vermutung, der Multicheck könnte aufgrund der Messung effektiver schulischer Fähigkeiten indirekt via schulische Probleme unentschuldigte Berufsschulabsenzen vorhersagen, bestätigt sich nicht. Die Signifikanz des Koeffizienten der Schulnoten, *zschnitt12*, hingegen bestätigt die Erwartung: Schulnoten können, möglicherweise aufgrund ihrer sozialen Dimension, durchaus zur Prognose von unentschuldigten schulischen Absenzen verwendet werden. Der gemessene Zusammenhang ist dabei beachtlich, wie Abbildung 12 aufzeigt: Sinkt die Durchschnittsnote der Bewerbungszeugnisse um einen Notenpunkt, steigt die Wahrscheinlichkeit unentschuldigter Absenzen um ungefähr 20 bis 25 Prozentpunkte. Die Abbildung illustriert ausserdem die Effekte derjenigen Kontrollvariablen, welche in den Modellen ebenfalls als signifikante Prädiktoren unentschuldigter Absenzen ausgewiesen werden – *stadt* und *sek*. Konkret verursachen Sekundarschüler wie auch Realschüler vom Land signifikant weniger unentschuldigte Absenzen als Realschüler aus der Stadt.

Um die Robustheit der Ergebnisse auf einflussreiche Beobachtungen zu überprüfen, wurde Pregibon's Beta ausgerechnet, das in Anlehnung an Cook's Distanz der linearen Regression den Einfluss einzelner Observationen auf das Regressionsergebnis approximiert (Long und Freese 2003). An-

[101] Die Informationskriterien wie auch McFadden's Pseudo-R^2 deuten dabei darauf hin, dass die Kantonsdummies zu integrieren sind. Einige von ihnen fallen allerdings weg, weil sie perfekt vorhergesagt werden. Deshalb verwenden die Modelle nur die Dummies für *be, lu, sg, zg* und *zh.*

[102] Die Parameter können dabei mit folgender Faustregel verglichen werden (Cameron und Trivedi 2009): $\hat{\beta}_{Logit} \cong 1.6 * \hat{\beta}_{Probit}$.

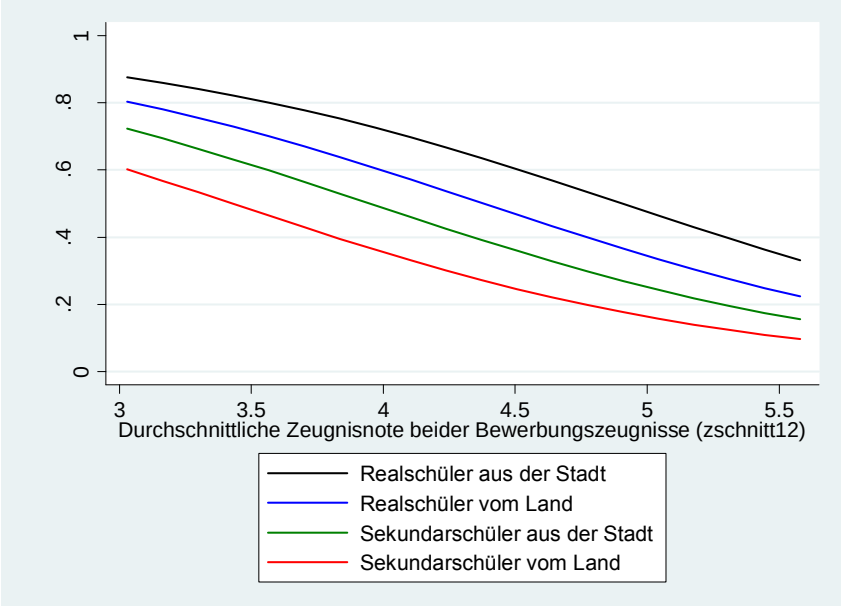
[103] Zur Beantwortung der Frage, welches Modell zu bevorzugen ist, können die Informationskriterien berücksichtigt werden: Beide, AIC und BIC, favorisieren das Logit-Modell leicht. Zudem lassen sich auch die Anteile an Observationen errechnen, welche vom Modell richtig vorhergesagt werden. Logit- und Probit-Modell erreichen dabei je 85.17% Richtige.

schliessend wurden jene Observationen weggelassen, welche die Ergebnisse stark beeinflussen. Die Ergebnisse für diese um 11 Beobachtungen kleinere Stichprobe sind ähnlich: Während *mcgesamt* insignifikant bleibt, werden die Effektgrössen der Zeugnisnoten (*zschnitt12*) und des Schultyps grösser. Erwartungsgemäss verändern sich allerdings vor allem die Koeffizienten der Regions- und Kantonsdummies, deren Punktschätzer zuvor wohl stark von den weggelassenen Observationen abhängten.[104] Im Übrigen sind die Resultate auch robust gegenüber der Berücksichtigung der sozialen Herkunft der Lernenden. Der Koeffizient von *berufe12* selbst ist im Modell insignifikant.

Weiter wurden die Ergebnisse auf die Möglichkeit hin überprüft, dass sich die Berufsschulen – ähnlich wie in der Notengebung – auch in der Handhabung unentschuldigter Absenzen unterscheiden. Tatsächlich zeigt ein Vergleich nach Berufsschulen, dass unentschuldigte Absenzen in einigen Berufsschulen oft, in anderen Berufsschulen sehr selten vorkommen: Der Anteil der Lernenden, welche unentschuldigte Absenzen zu verzeichnen haben, variiert in Berufsschulen mit mehr als 10 Migros-Lehrlingen zwischen 2.33 und 34.62%. Da erstaunt es nicht, dass Tests auf Unterschiedlichkeit der Varianz gemäss dem Verfahren von Baum (2006) feststellen, dass die standardisierten Pearson-Residuen der obigen Regressionen gruppenweise heteroskedastisch sind.

Einerseits wurden deshalb die Standardabweichungen von *mcgesamt* und *zschnitt12* der normalen Logit- und Probit-Modelle auch mithilfe eines Cluster-robusten Schätzers geschätzt. Die so erhaltenen Standardabweichungen sind in den eckigen Klammern in Tabelle 15 ausgewiesen (vgl. Long und Freese 2003). Dabei zeigt sich, dass der Cluster-robuste Schätzer für die zentralen Variablen eher kleinere Standardabweichungen liefert, die Differenzen zum normalen Schätzer allerdings gering sind.

ABBILDUNG 12: MIT DEM LOGIT-MODELL GESCHÄTZTE AUFTRETENSWAHRSCHEINLICHKEITEN UNENTSCHULDIGTER BERUFSSCHULABSENZEN IN DEN ERSTEN DREI LEHRSEMESTERN (LUNENTSCH01) NACH SCHULNOTEN, SCHULTYP UND RÄUMLICHER HERKUNFT



[104] Da sich die Koeffizienten der Kantons- und Regionen-Dummy-Variablen zudem je nach Modell und Spezifikation beträchtlich unterscheiden und stark von einzelnen positiven Ereignissen (d. h. *lunentsch01*=1) abhängen, wird auf eine Darstellung und eine Interpretation verzichtet.

Andererseits wurden die Regressionen auch mithilfe von Panel-Daten-Methoden durchgeführt. Konkret wurden je ein Random-Effects-Logit-, ein Random-Effects-Probit- und ein Fixed-Effects-Logit-Modell geschätzt (Rabe-Hesketh und Skrondal 2005; Wooldridge 2002a). Letzteres wird oft auch als konditionales Logit-Modell bezeichnet.[105] Die Ergebnisse dieser Robustheitsregressionen sind in Tabelle 33 in Anhang 7.K ausgewiesen. Zu berücksichtigen ist, dass die Regressionen auf einer etwas kleineren Stichprobe beruhen[106] und dass die Regionen-Dummy-Variablen für die Schätzung nicht verwendet werden konnten[107]. Trotzdem ist das Bild, das sich aus den Modellen ergibt, konsistent mit den bisherigen Ergebnissen: Lernende mit einem hohen Notendurchschnitt in den Bewerbungszeugnissen haben signifikant weniger unentschuldigte Absenzen in der Berufsschule als Lernende mit einem tiefen Notendurchschnitt. Dasselbe gilt für ehemalige Sekundarschüler im Vergleich zu ehemaligen Realschülern. Das Multicheck-Gesamtergebnis kann in diesem Kontext keinen prädiktiven Beitrag leisten – was im Übrigen auch für die einzelnen Testgebiete zutrifft.

4.5.2 LEHRVERTRAGSAUFLÖSUNGEN

Es mag als eine relativ unbedeutende Leistung der Volksschulnoten betrachtet werden, dass sich ihr Signalgehalt auch über die Prognose unentschuldigter Berufsschulabsenzen erstreckt. Doch sind unentschuldigte Absenzen nicht nur Ausdruck von schlechtem Sozialverhalten der Lernenden – wiederholte unentschuldigte Absenzen führen im schlimmsten Fall zu einer vorzeitigen Beendigung des Lehrverhältnisses (Moser et al. 2008). Entsprechend geben die Lehrlingsverantwortlichen der Migros vor allem disziplinarische Gründe an, die zu einer Lehrvertragsauflösung führen.

Eine Lehrvertragsauflösung ist die wohl unbestreitbarste Erfolgsvariable in der Berufsbildung. Kommt es zu einem Lehrabbruch, ist dies für beide Parteien eine unangenehme Angelegenheit: Während auf der Seite der Jugendlichen oft das Wohlbefinden beeinträchtigt wird und sich die Wiedereingliederung in den Arbeitsmarkt häufig schwierig gestaltet, ist die Auflösung für den Betrieb vor allem mit zeitlichem Aufwand und Kosten verbunden (Stalder und Schmid 2006). Dabei sind die Gründe einer vorzeitigen Lehrvertragsauflösung vielfältig (Bessey und Backes-Gellner 2008; Moser et al. 2008; Schmid und Storni 2004). Vielfach stehen aber gemäss Stalder und Schmid (2006) schulische Leistungsdefizite im Verbund mit anderen Problemen am Ursprung eines Lehrabbruchs.

Aus diesem Grund stellt sich im Rahmen dieser Arbeit die Frage, ob die Schulnoten des Bewerbungszeugnisses und der Detailhandels-Multicheck – möglicherweise indirekt – einen prognostischen Bei-

[105] Das konditionale FE-Modell zeichnet sich im Vergleich zu den RE-Modellen durch vorteilhafte statistische Eigenschaften aus. Insbesondere ist es weniger anfällig auf Missspezifikationen und unterstellt im Gegensatz zu den RE-Modellen keine Verteilung des schulspezifischen Effekts, weil es mithilfe eines konditionalen Maximum-Likelihood-Verfahren geschätzt wird (Rabe-Hesketh und Skrondal 2005).

[106] Im FE-Modell werden – im Gegensatz zum RE-Modell – zumindest theoretisch ähnlich wie in einer Dummy-Variablen-Regression einzelne Dummy-Koeffizienten für jede Berufsschule geschätzt. Da in der Stichprobe in drei berücksichtigten Berufsschulen allerdings keine unentschuldigten Absenzen vorkommen und damit auch keine Variation in der exogenen „Dummy"-Variable auftritt, können 29 Observationen dieser Berufsschulen nicht berücksichtigt werden. Zudem wurden zwei Ausreisser der Stichprobe entfernt. Die Observationen fielen auf, weil sie vom Modell nicht gut vorhergesagt wurden bzw. sehr grosse standardisierte Pearson-Residuen aufwiesen (vgl. Long und Freese 2003).

[107] Die beiden Regionsdummies mussten weggelassen werden, weil das iterative Verfahren der clogit-Funktion zur Schätzung eines Fixed-Effects-Logit-Modells in Stata beim Einbezug der Regionendummies jeweils daran scheiterte, die Log-Likelihood-Funktion zu minimieren.

trag zu leisten vermögen, wenn es darum geht, die Wahrscheinlichkeit eines Lehrabbruches vorherzusagen. Dementsprechend wurde angestrebt, auch die Daten derjenigen Jugendlichen (mit Lehrbeginn 2007 und 2008) zu erheben, die von einer Lehrvertragsauflösung betroffen waren. Leider erwies sich dieses Vorhaben nur als bedingt umsetzbar: Einerseits konnten die Daten der betroffenen Lehrlinge der Migros Luzern aus Datenschutzgründen nicht erhoben werden. Deshalb werden die Lernenden der Migros Luzern im Folgenden ausgeklammert. Andererseits wurde die systematische Erfassung auch deshalb eingeschränkt, weil die Lehrlingsausbildner der Region Aare bereits damit begonnen hatten, den ehemaligen Lehrlingen die Bewerbungsdossiers – und damit die Multicheck-Auswertungen – zurückzuschicken. Insgesamt umfasst die Stichprobe deshalb 14 Lehrabbrüche auf 250 Lernende, was einer Auflösungsquote von 5.6% entspricht. Diese Quote ist zwar ähnlich mit jener, die in Mühlemann et al. (2007) ausgewiesen ist, liegt allerdings klar unter den 19-30% im Detailhandel, welche Stalder und Schmid (2006) für die Jahre 1995 bis 2002 im Kanton Bern finden.

Trotz der Eingeschränktheit der Stichprobe wurde die binäre Variable *labbruch* analog des vorigen Abschnitts nun innerhalb je eines Probit- und Logit-Modelles untersucht.[108] Dieses Mal wurden die exogenen Variablen explizit anhand ihrer Signifikanz ausgewählt. Damit soll vermieden werden, dass das Modell, das bereits stark von einzelnen positiven *Outcomes* – das heisst *labbruch=1* – abhängen dürfte, auch noch aufgrund des Verlusts an Freiheitsgraden ineffizient geschätzt wird.[109]

Tabelle 16 zeigt die Schätzresultate für die gebildeten Modelle. Dabei deuten die Informationskriterien wie auch McFadden's Pseudo-R^2 darauf hin, dass das Probit- dem Logit-Modell leicht überlegen

TABELLE 16: PROBIT- UND LOGIT-MODELL DER WAHRSCHEINLICHKEIT EINER LEHRVERTRAGSAUFLÖSUNG

	labbruch	
	(1)	(2)
VARIABLEN	Probit	Logit
mcgesamt	0.0212	0.0387
	(0.0173)	(0.0345)
zschnitt12	-0.632+	-1.138+
	(0.332)	(0.634)
weiblich	0.325	0.760
	(0.301)	(0.630)
muttersp	0.191	0.375
	(0.311)	(0.641)
stadt	0.315	0.666
	(0.325)	(0.649)
sek	-1.148**	-2.562*
	(0.443)	(1.091)
Konstante	-0.281	-0.597
Observationen	250	250
Pseudo-R-Quadrat	0.116	0.113

Kantons- und Regionendummies nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

[108] Da die Eintretenswahrscheinlichkeit von 5.6% wiederum auf eine Ecklösung hindeutet, wurde erneut darauf verzichtet, sich auf eine der beiden Verteilungen festzulegen (vgl. Abschnitt 4.5.1).

[109] Folgende Variablen des Basismodells wurden deshalb aufgrund von Parametersignifikanz und Modell-Kennzahlen (McFadden's R^2, AIC und BIC) weggelassen: *laltera, lalterj,* und *zuerich.*

ist. Wichtiger ist allerdings der Befund, dass der Durchschnitt der Bewerbungszeugnisse *zschnitt12* zumindest auf dem 10%-Signifikanzniveau tatsächlich negativ mit der Wahrscheinlichkeit zusammenhängt, einen Lehrabbruch zu verzeichnen. Dies ist auch deshalb bemerkenswert, weil die Vertragsauflösungen gemäss den Vertragsauflösungsschreiben nur in vier Fällen tatsächlich aufgrund von ungenügenden schulischen Leistungen zustande kamen.[110] Neben dem Koeffizienten für *zschnitt12* ist nur noch jener des Schultyps signifikant. Dieses Ergebnis kommt zustande, weil von den 14 erfassten Lehrabbrüchen nur ein einziger auf einen Sekundarschüler fiel.[111] Insofern ist diese Erkenntnis nicht robust. Gleichzeitig implizieren die Ergebnisse des Modells, dass im vorliegenden Fall die Wahrscheinlichkeit einer Lehrvertragsauflösung nicht von Geschlecht, Muttersprache, räumlicher Herkunft oder – für die Fragestellung entscheidend – vom Multicheck-Gesamtergebnis abhängt. Somit bestätigt sich der geringe Signalwert des Tests auch in diesem Kontext. Sollte sich dieses Ergebnis in einer umfassenderen Erhebung bestätigen, ist es äusserst fragwürdig, inwiefern der Test tatsächlich die „Berufseignung" eines Berufsinteressierten misst.

Allerdings dürfen die hier präsentierten Ergebnisse nicht als robust bezeichnet werden, weil sie auf einer kleinen Stichprobe von Lehrabbrechern beruhen. Trotzdem liefern die Resultate einige Hinweise, dass das Signal Volksschulnote zur Vorhersage der Wahrscheinlichkeit einer Lehrvertragsauflösung genutzt werden kann. Die aus der Theorie abgeleitete Vermutung scheint sich also zu bestätigen. Der Zusammenhang ergibt sich möglicherweise indirekt aufgrund der statistisch signifikanten Beziehung zwischen Schulnoten, Berufsschulnoten und der Wahrscheinlichkeit unentschuldigter Absenzen. Vielleicht ist er aber eine direkte Auswirkung der sozialen Dimension von Schulnoten.

4.6 Externe Validität der Ergebnisse

Die Erkenntnisse, welche sich aus den aus den Daten gewinnen lassen, liefern ein konsistentes Bild: Während der Signalgehalt des Detailhandels-Multichecks gesamthaft relativ klein ausfällt und durch die Berücksichtigung kontextueller Variablen deutlich eingeschränkt wird, erweisen sich Schulnoten als valide Prädiktoren der betrachteten abhängigen Variablen. Neben dem Problem, dass diese Erkenntnisse möglicherweise vor allem beim Multicheck-Gesamtergebnis aufgrund der Linkszensiertheit gegen unten verzerrt sind – eine Sache der internen Validität –, stellt sich auch die Frage, inwiefern die Ergebnisse extern valide, das heisst „über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinausgehend generalisierbar" (Bortz und Döring 2006, 53) sind. Die hohe Standardisierung der theoretischen Ausbildung der Detailhandelslehre und die Tatsache, dass es sich bei der vorliegenden Studie um eine Vollerhebung handelt, die zudem mit multivariaten Verfahren Kontexteinflüsse wie sozioökonomische und räumliche Herkunft oder persönliche Merkmale kontrolliert, wirken sich positiv auf die externe Validität der Fallstudie aus.

Gleichwohl begrenzen gewisse Faktoren die Generalisierbarkeit der Erkenntnisse. Insbesondere mindert die Beschränkung auf einen einzelnen Grossbetrieb die externe Validität wohl deutlich.

[110] Häufigster Kündigungsgrund sind ungenügende betriebliche Leistungen. Weiter werden das Verhalten am Arbeitsplatz, unentschuldigte Absenzen, falsche Berufswahl und Motivationsprobleme angegeben.

[111] Die Punktschätzer der anderen Koeffizienten hängen allerdings nicht wesentlich von dieser einzelnen Observation ab. Wird sie weg gelassen, werden die Koeffizienten – trotz des Verlusts aller Sekundarschüler in der Stichprobe – nur unwesentlich anders geschätzt.

Zwar ist die Lehrsituation der einzelnen Lernenden weniger durch die übergeordnete Einheit – die Genossenschaft –, als durch den einzelnen Ausbildungsbetrieb – die Migros-Niederlassung – geprägt, wodurch die Daten durchaus Varianz in der kontextuellen Bedingtheit der Ausbildungskonstellation enthalten. Doch zeigen die Ergebnisse von Moser (2004c), dass sich (Gross-) Betriebe in nicht zu unterschätzenden Ausmass in ihrer Lehrlingsselektion unterscheiden, während Klein- im Vergleich zum hier betrachteten Grossbetrieb ohnehin eine andere Selektionspraxis und Lehrlingsausbildung haben (Imdorf 2007a). Ausserdem belegen Studien einen relativ grossen betrieblichen Einfluss auf die Leistungsentwicklung der Lehrlinge (Stamm et al. 2009) und die Wahrscheinlichkeit einer Lehrvertragsauflösung (Bessey und Backes-Gellner 2008). Diese Unterschiede rühren beispielsweise von verschiedenartigen Förderstrategien der Betriebe her: Während die einen die schulisch schwächeren Lernenden unterstützen, versuchen andere Betriebe die Leistungsstärkeren zu fördern (Mühlemann et al. 2007). Zwar arbeiten die drei Migros-Genossenschaften den Aussagen der Lehrlingsbetreuer nach zu urteilen mit teilweise unterschiedlichen Selektionsphilosophien und der Einfluss des Lehrbetriebes auf die schulischen Ergebnisse ist wohl relativ gering, da mit den Noten aus dem ersten und dritten Semester frühe berufsschulische Leistungsausweise betrachtet wurden. Trotzdem ist schon nur aufgrund des betriebsspezifischen Selektionsverhaltens nicht gesichert, dass die Ergebnisse in anderen Betrieben des Schweizer Detailhandels replizierbar wären.

Einen ersten Hinweis darauf liefert die von Multicheck selbst erarbeitete Studie zum prädiktiven Gehalt des Detailhandels-Multichecks: Die Studie weist einen Korrelationskoeffizienten zwischen Lehrabschlussnoten und Multicheck von 0.4 aus, was deutlich über dem in dieser Studie gefundenen Zusammenhang liegt. Die Unterschiede in den Ergebnissen treten auf, obwohl die Stichprobe der Multicheck-Studie aus Lernenden der Migros und der Coop besteht und der Kontext der Studien – ausser der betrachteten Zeitspanne – sehr ähnlich ist. Die Autoren der Studie verweisen denn auch darauf, dass „bei Detailauswertungen der einzelnen Firmen sehr grosse Schwankungen festgestellt werden“ (Multicheck 2008, 3). Dies zeige einen grossen Einfluss der Ausbildungsverantwortlichen.

Schon allein wegen der betrieblichen Bedingtheit der Ergebnisse, aber auch wegen der Komplexität der beruflichen Bildung in der Schweiz (Moser 2004c), sind Schlüsse, welche die Erkenntnisse dieser Studie über den Zusammenhang der Detailhandelslehre hinaus verallgemeinern, also vorsichtig zu formulieren. Schliesslich könnten auch die Ergebnisse zum Signalgehalt der Schulnoten und Multicheck relativ stark kontextabhängig sein – ein Fakt, den auch Baron-Boldt (1989, 26 f.) in ihrer Metaanalyse konstatierte. Ihre Studie verdeutliche, „dass Untersuchungen über die Validität von Schulabschlussnoten zur Prognose von Ausbildungs- und Studienerfolg zu sehr unterschiedlichen Ergebnissen gelangen.“ Ähnlich könnte es mit dem Multicheck verhalten, der für gewisse Berufe oder in anderen Versionen (z. B. Multicheck Kaufmann) durchaus eine grössere Aussagekraft aufweisen könnte. Insgesamt sollte aber trotz der nicht von der Hand zu weisenden Eingeschränktheit der externen Validität der Ergebnisse nicht vergessen werden, dass die Migros in der Detailhandelsbranche der Schweiz ein äusserst bedeutender Arbeitgeber ist, und damit die Legitimität und Bedeutung der Ergebnisse gerade bezüglich Multicheck von durchaus branchenweiter Tragweite sind.

5 SCHLUSSFOLGERUNGEN

Eines zeigen die Daten deutlich, die für die hier vorliegende, quantitative Fallstudie in einer Vollerhebung von über 300 Lernenden der drei Migros-Genossenschaften Aare, Luzern und Zürich erhoben wurden: Der populäre Berufseignungstest „Multicheck Detailhandel und Service" vermag die Hoffnungen, die man aus Signalling-theoretischer Sicht in ihn stecken kann, nicht zu erfüllen. Der Signalwert des Berufseignungstests für die Migros, gemessen an seiner prädiktiven Validität in Bezug auf die drei betrachteten Lehrerfolgskenngrössen, ist klein: Weder ist er dazu in der Lage, einen Beitrag zur Vorhersage der Wahrscheinlichkeit unentschuldigter Berufsschulabsenzen zu leisten, noch hängen gute Ergebnisse im Detailhandels-Multicheck statistisch mit der Wahrscheinlichkeit einer Lehrvertragsauflösung zusammen. Schliesslich trägt der Test auch wenig dazu bei, die Noten vorherzusagen, welche die Lernenden im ersten und dritten Semester der Berufsschule erzielen. Der prognostische Mehrwert, den der Multicheck in die verwendeten multivariaten Regressionsmodelle unter Kontrolle kontextueller Einflüsse wie Geschlecht, schulischer Laufbahn oder sozialer und räumlicher Herkunft einbringt, ist gering.

Insgesamt verbessert der Detailhandels-Multicheck die Lehrlingsauswahl der Migros nur wenig – denn die Informationen, die der Multicheck dem Betrieb liefert, sind offenbar insbesondere in den schulischen Leistungsausweisen – Schulnoten und auf Sekundarstufe I besuchter Schultyp – bereits enthalten. Vorausgesetzt die Selektionsverantwortlichen können die schulischen Signale berufsschulischer Leistungen richtig interpretieren, leistet der Test für die Migros also nur einen kleinen Beitrag, die Kompetenzen der Schulabgänger besser zu signalisieren.

Diese Befunde entsprechen nicht den Erwartungen, da der Multicheck vorgibt, für den Beruf relevante, effektive schulische Kompetenzen und kognitive Fähigkeiten der Berufsinteressierten in standardisierter und damit vergleichbarer Weise zu erheben. Gemäss der ökonomischen und psychologischen Literatur müsste ihm dies zu einem beachtlichen, eigenständigen Signalwert verhelfen. Wie ist das unerwartet schlechte Abschneiden des Tests zu erklären?

Es gibt viele mögliche Erklärungsansätze. Die wichtigsten haben aber mit Einschränkungen der internen und externen Validität dieser Arbeit und mit der Konzeption des Multichecks zu tun. So begrenzen kleine Fallzahlen in einigen Regressionsmodellen die interne Validität der Ergebnisse. Daneben dürfte vor allem ins Gewicht fallen, dass die Stichprobe einer Vorselektion unterliegt. Denn die Migros misst dem Multicheck als Selektionsinstrument gemäss den Daten viel Gewicht bei, so dass Lernende mit ungenügenden Multicheck-Ergebnissen bei den betrachteten Migros-Genossenschaften nur mit sehr geringer Wahrscheinlichkeit eine Lehrstelle erhalten. Deshalb kann der Signalgehalt des Tests mit den vorliegenden Daten nur über die oberen beiden Drittel seiner eigenen Skala bewertet werden. Dies führt dazu, dass der Signalwert in den gebildeten Modellen tendenziell unterschätzt wird. Schliesslich ist die Arbeit als Fallstudie konzipiert, was die externe Validität der Resultate mindert. So dürften die Ergebnisse nur sehr bedingt über den Kontext des Unternehmens, des betrachteten Berufs und der analysierten Multicheck-Version hinaus verallgemeinerbar sein.

Diese Arbeit liefert aber auch Evidenz, dass der tiefe Signalgehalt des Multichecks teilweise konzeptionellen Mängeln geschuldet ist. Es scheint, als ob der Multicheck die *relevanten* schulischen und kognitiven Fähigkeiten nicht oder nur fehlerhaft misst. So zeigt ein genauer Blick auf die Signalwerte einzelner Multicheck-Testgebiete, dass diese für sich genommen prognostisch valider sind als ihr Aggregat – das Gesamtergebnis des Tests. Ausserdem können einzelne Multicheck-Testgebiete zur Prognose inhaltlich entsprechender Berufsschulfächer verwendet werden. Andere Testgebiete wiederum weisen überhaupt keinen Signalgehalt auf, und einige wichtige Berufsschulnoten vermag der Test fast gar nicht zu prognostizieren. So liegt die Vermutung nahe, dass der Test auch deshalb einen geringen Signalgehalt aufweist, weil er den inhaltlichen Anforderungen der Detailhandelslehre zu wenig entspricht.

Übertragen auf die Messfehlermodelle der Ökonomen (De Paola und Scoppa 2010; Grant 2007; Schwerdt und Wössmann 2009) heisst dies: Die Fehler im Produktivitätssignal, die durch die Einmaligkeit der externen Leistungsmessung entstehen, scheinen grösser zu sein, als die Fehler, die auf *Noise* in der Leistungsmessung der Volksschulnoten basieren. So belegt diese Arbeit – bis zu einem gewissen Grad kontraintuitiv und in Kontrast zur breiten betrieblichen Kritik an den Volksschulnoten – dass der standardisierte Berufseignungstest den Schulzeugnissen in Bezug auf die Prognose aller drei betrachteten Kenngrössen des Lehrerfolgs unterlegen ist.

Die Schulnoten sind, zumindest im Sinne einer betrieblichen Verwertbarkeit für die Lehrlingsauswahl in der Migros, kein „sinnentleertes Leistungszertifikat" (Kronig 2007, 201): Gute Schulnoten oder auch ein höheres besuchtes Schulniveau signalisieren eine signifikant geringere Wahrscheinlichkeit unentschuldigter berufsschulischer Absenzen und eine kleinere Wahrscheinlichkeit einer Lehrvertragsauflösung. Ausserdem leisten Schulnoten und Schultyp über alle geschätzten Modelle, Spezifikationen und Schätzverfahren hinweg einen Beitrag zur Prognose der Erst- und Drittsemester-Berufsschulnoten. Der Umfang des Beitrags scheint dabei vom kantonalen Schulsystem und von der Muttersprache abzuhängen: Zeugnisse aus den Kantonen Bern und Aargau und von Jugendlicher deutscher Muttersprache sind aussagekräftiger. Ob das Zeugnis von einem Real- oder von einem Sekundarschüler stammt, ist hingegen nicht von Bedeutung. Der auf Sekundarstufe I besuchte Schultyp schlägt sich aber in einem Niveaueffekt nieder: Einstige Realschüler sind bei sonst identischen Merkmalen und gleichen Schulnoten in der Berufsschule um gut eine Viertelnote schlechter als ehemalige Sekundarschüler. Zusammen erklären Schulniveau und Schulnoten so in etwa denselben Anteil der Varianz in den Berufsschulnoten, welche die Regressionsmodelle aufklären, wie alle anderen Kontrollvariablen zusammen.

Doch wenn Schulnoten und Schulniveauzuteilung systematisch von Lehrer, Schulklasse, Herkunft, Eltern und Zufälligkeiten abhängen (Kronig 2007), wie erklärt sich der im Vergleich zum standardisierten Berufseignungstest grosse Signalwert der schulischen Leistungsausweise? Es ist angesichts der Menge an gegenteiligen Forschungsbefunden unwahrscheinlich, dass es hauptsächlich die Messung effektiver schulischer Fähigkeiten ist, die den Signalwert der Volksschulnoten bestimmt. Dies zeigt sich auch in den geschätzten Regressionsmodellen: Der Koeffizient der Volksschulnoten bleibt beinahe unberührt, wenn das Multicheck-Gesamtergebnis oder verschiedene, schulische Multicheck-

Testgebiete in die Modelle hinzugefügt werden. Unter der Voraussetzung, dass der Multicheck wirklich effektive schulische Fähigkeiten ausweist, haben diese mit den Schulnoten nur wenig gemein.
Vielmehr sind im Kontext dieser Arbeit wohl drei andere Faktoren ursächlich für den Signalwert der Schulnoten: Erstens profitiert der Signalgehalt der Schulnoten von der Reduktion in der lehrerabhängigen Fehlerkomponente, welche die Mittelwertbildung hervorruft (vgl. Grant 2007). Denn obwohl in dieser Arbeit auch Einzelnoten eine signifikante prognostische Aussagekraft in Bezug auf die Berufsschulnoten aufweisen, erreicht der Notenschnitt doch einen deutlich höheren und robusteren Signalgehalt. Zweitens dürfte dem Signalwert der Volksschulnoten die wiederholte Leistungsmessung, die ihnen zugrunde liegt, zuträglich sein, da Noten damit weniger anfällig auf einzelne, von der Tagesform abhängige Messfehler sind, die beim Multicheck eine Rolle spielen könnten (vgl. De Paola und Scoppa 2010). Und drittens wurde in Einklang mit der Forschungsliteratur die Hypothese aufgestellt, dass Schulnoten, indem sie Eigenschaften wie Freundlichkeit, Offenheit und Motivation signalisieren, eine soziale Dimension aufweisen, die ihren Signalgehalt erhöht.
Die Frage, wie der Signalwert eines derart gescholtenen Leistungszertifikats, wie es die schulischen Leistungsausweise sind, zustande kommt, bedarf aber weiterer Forschungsaufmerksamkeit – sollte sich der robuste, eigenständige Signalgehalt der Schulnoten in anderen Kontexten bestätigen. Es werden Studien benötigt, die den Informationsgehalt und die Bedingtheiten des Signalwerts von Schulnoten umfassend modellieren, da es sich um ein komplexes Signalling-Problem handelt, in dem sich ein Betrieb bei der Interpretation eines Schulzeugnisses in der Lehrlingsauswahl wiederfindet. Solche Studien würden helfen, die betriebliche Verwertung von Schulzeugnissen besser zu verstehen, könnten dazu beitragen, das Signalling zwischen Schule und Berufsschule zu begreifen und darauf aufbauend, es zu verbessern, indem sie möglicherweise wohlfahrtssteigernde Massnahmen in der Gestaltung eines Schulzeugnisses offenbaren würden (Bishop 1994; Costrell 1994). So liesse sich beispielsweise unter entsprechender Berücksichtigung nichtkognitiver Fähigkeiten in Regressionsmodellen die Hypothese überprüfen, inwiefern der Schulnotensignalgehalt durch die soziale Dimension der Schulnoten bestimmt wird.
Ähnlich zentrale Forschungsaufmerksamkeit sollte den in der Schweiz eingesetzten Berufseignungstests zukommen, zumal diese die Bildungschancen eines Jugendlichen heutzutage stark determinieren. So stellt sich die Frage, ob sich die ernüchternden Resultate in Bezug auf den Detailhandels-Multicheck auf einen überbetrieblichen Rahmen und auf die anderen Multicheck-Versionen und Eignungstests übertragen lassen. Wie sollten die Betriebe die Eignungstests in ihrer Selektion einsetzen? Wie liesse sich der Signalwert des Tests selbst verbessern?
Denn Optimierungspotential gibt es allemal. Die Daten zeigen, dass auch sehr gute Multicheck-Ergebnisse kein Garant guter berufsschulischer Leistungen kurz nach Lehrantritt sind, und, umgekehrt, dass auch ein schlechtes Abschneiden im Multicheck sehr gute theoretische Lehrleistungen nicht ausschliesst. So sind die bereits für den „basic-check“ geäusserten Zweifel (Jungo und Zihlmann 2002) über die Berechtigung der bedeutsamen Diagnose, welche der Eignungstest über die „berufsbezogene Eignung“ der Jugendlichen stellt, wohl auch für den Multicheck angebracht – die Grenzen von 40 und 60 Punkten im Gesamtergebnis, die Multicheck ausweist, erscheinen arbiträr.

Zweifel an der Diagnose sind auch deshalb berechtigt, (1) weil es für die Migros mehr Sinn macht, statt auf das Gesamtresultat des Tests nur auf das Englisch- und Merkfähigkeitsresultat zu schauen, wenn sie die durchschnittlichen Lehrleistungen eines Bewerbers prognostizieren will; (2) weil die Migros selektiv nur einzelne Testresultate zur Beurteilung entsprechender schulischer Kompetenzen ihrer Bewerber einsetzen sollte; (3) weil die Migros den kognitiven Teil des Tests mit Ausnahme der Merkfähigkeit für die Vorhersage berufsschulischer Leistungen schlicht ignorieren sollte; und (4) weil Volksschulnoten der Migros die Berufseignung wesentlich robuster und aussagekräftiger signalisieren als der Multicheck. Schliesslich sei die Frage erlaubt – die sich im Übrigen auch auf die anderen Multicheckversionen übertragen lässt: Wie kann ein Test, der mich als Kandidat schlussendlich in Prozentrangwerte einteilt – mir also sagt, wie gross der Anteil der Kandidaten ist, die schlechter, und wie gross der Anteil der Kandidaten ist, die besser als ich abgeschnitten haben – wie kann ein Test bei einem solchen Aggregationsverfahren sicherstellen, dass ich mich exakt dann nicht mehr für einen Beruf eigne, wenn ich schlechter als 60 Prozent der anderen Absolventen bin?

6 LITERATURVERZEICHNIS

6.1 LITERATUR

AI, C., und NORTON, E. C., (2003). "Interaction terms in logit and probit models." *Economics Letters*, 80(1), 123-129.

AMOS, J., AMSLER, F., MARTIN-JAHNCKE, M., und 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.

ASHTON, D. N., und MAGUIRE, M. J., (1980). "The function of academic and non-academic criteria in employers' selection strategies." *British Journal of Guidance & Counselling*, 8(2), 146-157.

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.

BAUM, C. F., (2006). "Stata tip 38: testing for groupwise heteroskedasticity." *Stata Journal*, 6(4), 590-592.

BBT, (2004). *Verordnung über die berufliche Grundbildung vom 8. Dezember 2004*, Bern: Bundesamt für Berufsbildung und Technologie (BBT).

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

BESSEY, D., und BACKES-GELLNER, U., (2007). "Premature apprenticeship terminations: an economic analysis." *ISU Economics of Education Working Paper Series 2*.

BESSEY, D., und BACKES-GELLNER, U., (2008). "Warum Jugendliche eine Ausbildung abbrechen." *Panorama*, 23(Sonderheft: Leading House), 20 f.

BETTS, J. R., und MORRELL, D., (1999). "The determinants of undergraduate grade point average." *Journal of Human Resources*, 34(2), 268-293.

BFS, (2005a). *PISA 2003 – Einflussfaktoren auf die kantonalen Ergebnisse*. BFS, Neuenburg.

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

BFS, (2009a). *PISA 2006: Analysen zum Kompetenzbereich Naturwissenschaften. Rolle des Unterrichts, Determinanten der Berufswahl, Vergleich von Kompetenzmodellen*. BFS, Neuenburg.

BFS, (2009b). *PISA 2006: Kompetenzen für das Leben – Schwerpunkt Naturwissenschaften. Nationaler Bericht*. BFS, Neuenburg.

BILLS, D. B., (2003). "Credentials, signals, and screens: explaining the relationship between schooling and job assignment." *Review of Educational Research*, 73(4), 441-469.

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

BISHOP, J. H., (1997). "The effect of national standards and curriculum-based exams on achievement." *The American Economic Review*, 87(2), 260-264.

BISHOP, J. H., und WÖßMANN, L., (2004). "Institutional effects in a simple model of educational production." *Education Economics*, 12(1), 17-38.

BOES, S., und WINKELMANN, R., (2006). "Ordered response models." *Allgemeines Statistisches Archiv*, 90(1), 167-181.

BORTZ, J., und DÖRING, N., (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4. Auflage, Heidelberg: Springer Medizin Verlag.

BUIS, M. L., (2010). "Stata tip 87: Interpretation of interactions in nonlinear models." *Stata Journal*, 10(2), 305-308.

CAMERON, A. C., GELBACH, J. B., und MILLER, D. L., (2008). "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics*, 90(3), 414-427.

CAMERON, C. A., und TRIVEDI, P. K., (2009). *Microeconometrics using Stata*, College Station: StataCorp.

CHAN, W., HAO, L., und SUEN, W., (2007). "A signaling theory of grade inflation." *International Economic Review*, 48(3), 1065-1090.

COHN, E., COHN, S., BALCH, D. C., und BRADLEY, J., (2004). "Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank." *Economics of Education Review*, 23(6), 577-586.

CONSTANT, A., und KONSTANTOPOULOS, S., (2003). "School effects and labor market outcomes for young adults in the 1980s and 1990s." *Applied Economics Quarterly*, 49(1), 5-22.

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

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

DIEKMANN, A., (2005). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*, 13. Auflage, Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.

DUCKWORTH, A. L., und SELIGMAN, M. E. P., (2006). "Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores." *Journal of Educational Psychology*, 98(1), 198-208.

ECKSTEIN, P. P., (2008). *Angewandte Statistik mit SPSS*, 8. Auflage, Wiesbaden: Gabler Verlag.

FEISCHMANN, D., (2004). "Was Leistung heisst, soll wieder die Schule sagen." *Panorama*, 17(6), 6-7.

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

GALLIKER, R., (2003). "Tests als Hilfsmittel für die Lehrlingsselektion." *bbaktuell*, 31(103), 1-3.

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

GREENE, W. H., (2003). *Econometric analysis*, 5. Auflage, Upper Saddle River: Prentice Hall.

HÄBERLIN, U., IMDORF, C., und 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.

HÄFELI, K., und SCHELLENBERG, C., (2009). *Erfolgsfaktoren in der Berufsbildung bei gefährdeten Jugendlichen*, Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektion (EDK).

HANUSHEK, E. A., KAIN, J. F., MARKMAN, J. M., und RIVKIN, S. G., (2003). "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18(5), 527-544.

HANUSHEK, E. A., und WÖSSMANN, L., (2006). "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries." *The Economic Journal*, 116(510), C63-C76.

HECKMAN, J., und 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.

HECKMAN, JAMES J., STIXRUD, J., und URZUA, S., (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24(3), 411-482.

HELL, B., TRAPMANN, S., und SCHULER, H., (2008). "Synopse der Hohenheimer Metaanalysen zur Prognostizierbarkeit des Studienerfolgs und Implikationen für die Auswahl- und Beratungspraxis", in H. Schuler und B. Hell, (Hrsg.). Göttingen: Hogrefe, 43-54.

HERZOG, W., NEUENSCHWANDER, M. P., und WANNACK, E., (2004). "Wie verlaufen Berufswahlprozesse?" *Panorama*, 17(2), 36 f.

HERZOG, W., NEUENSCHWANDER, M. P., und WANNACK, E., (2006). *Berufswahlprozess – wie sich Jugendliche auf ihren Beruf vorbereiten*, Bern: Haupt.

HUPKA, S., SACCHI, S., und STALDER, B. E., (2006). *Herkunft oder Leistung? Analyse des Eintritts in eine zertifizierende nachobligatorische Ausbildung anhand der Daten des Jugendlängsschnitts TREE*. TREE, Bern.

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

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

IMDORF, C., (2007b). "Warum Lehrbetriebe ausländische Jugendliche meiden." *Panorama*, 20(2), 29 f.

IMDORF, C., (2007c). "Weshalb ausländische Jugendliche besonders grosse Probleme haben, eine Lehrstelle zu finden", in H.-U. Grunder und L. von Mandach, (Hrsg.), *Auswählen und ausgewählt werden. Integration und Ausschluss von Jugendlichen und jungen Erwachsenen in Schule und Beruf*. Zürich: Seismo Verlag, 100-111.

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

INGENKAMP, K., (1995). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte*, Weinheim: Beltz.

JOHNSON, V. E., (2003). *Grade inflation. A crisis in college education*, New York: Springer-Verlag.

JUNGO, D., und ZIHLMANN, R., (2002). "Empirische Untersuchung des basic-check." *Panorama*, 15(2), 30 f.

KÉZDI, G., (2004). "Robust standard error estimation in fixed-effects panel models." *Hungarian Statistical Review*, 9(Special English Issue), 95-116.

KRAMER, J., (2009). *Metaanalytische Studien zu Intelligenz und Berufsleistung in Deutschland*, Rheinische Friedrich-Wilhelms-Universität zu Bonn, Bonn.

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

LAIDRA, K., PULLMANN, H., und ALLIK, J., (2007). "Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school." *Personality and Individual Differences*, 42(3), 441-451.

LEITER, J., und BROWN, J. S., (1985). "Determinants of elementary school grading." *Sociology of Education*, 58(3), 166-180.

LEKHOLM, A. K., und 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.

LIENERT, G. A., und RAATZ, U., (1998). *Testaufbau und Testanalyse*, 6. Auflage, Weinheim: Psychologie VerlagsUnion.

LONG, J. S., und FREESE, J., (2003). *Regression models for categorical dependent variables using Stata*, College Station: Stata Press.

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

MEYER, T., (2009). "Wer hat, dem wird gegeben: Bildungsungleichheit in der Schweiz", in C. Suter, S. Perrenoud, R. Levy, U. Kuhn, D. Joye, und P. Gazareth, (Hrsg.), *Sozialbericht 2008. Die Schweiz vermessen und verglichen*. Zürich: Seismo, 60-81.

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

MILLER, S. R., und ROSENBAUM, J. E., (1997). "Hiring in a Hobbesian world." *Work and Occupations*, 24(4), 498-523.

MILLMAN, J., SLOVACEK, S. P., KULICK, E., und MITCHELL, K. J., (1983). "Does grade inflation affect the reliability of grades?" *Research in Higher Education*, 19(4), 423-429.

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

MOSER, U., (2004a). "Deutsch und Mathematik bleiben der Schlüssel zum Erfolg." *Panorama*, 17(1), 46 f.

MOSER, U., (2004b). *Jugendliche zwischen Schule und Berufsbildung. Eine Evaluation bei Schweizer Grossunternehmen unter Berücksichtigung des internationalen Schulleistungsvergleichs PISA*. NFPNR 43, Bern, Aarau.

MOSER, U., (2004c). *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., und WÜEST, A., (2007). *Lehrlingsausbildung – ökonomisch betrachtet. Ergebnisse der zweiten Kosten-Nutzen-Studie*, Zürich, Chur: Rüegger Verlag.

MURNANE, R. J., WILLETT, J. B., BRAATZ, M. J., und 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.

MURNANE, R. J., WILLETT, J. B., DUHALDEBORDE, Y., und TYLER, J. H., (2000). "How important are the cognitive skills of teenagers in predicting subsequent earnings?" *Journal of Policy Analysis and Management*, 19(4), 547-568.

NEUENSCHWANDER, M. P., und MALTI, T., (2009). "Selektionsprozesse beim Übergang in die Sekundarstufe I und II." *Zeitschrift für Erziehungswissenschaft*, 12(2), 216-232.

NEUENSCHWANDER, M. P., und WISMER, N., (2010). "Wichtige Rolle der überfachlichen Kompetenzen." *Panorama*, 23(1), 16-17.

PERRIG-CHIELLO, P., PERRIG, W. J., und BUSER, N., (1997). "Ist weiblicher Schulerfolg wirklich besser vorhersagbar als männlicher? Geschlechtsspezifische Prognostizierbarkeit von Maturitätsnoten." *Bildungsforschung und Bildungspraxis*, 19(2), 174-187.

POROPAT, A. E., (2009). "A meta-analysis of the five-factor model of personality and academic performance." *Psychological Bulletin*, 135(2), 322-338.

RABE-HESKETH, S., und SKRONDAL, A., (2005). *Multilevel and longitudinal modeling using Stata*, College Station: StataCorp.

RESH, N., (2009). "Justice in grades allocation: teachers' perspective." *Social Psychology of Education*, 12(3), 315-325.

ROTH, P. L., BEVIER, C. A., SWITZER, F. S., und SCHIPPMANN, J. S., (1996). "Meta-analyzing the relationship between grades and job performance." *Journal of Applied Psychology*, 81(5), 548-556.

SCHLOTTER, M., SCHWERDT, G., und WÖSSMANN, L., (2009). "Econometric methods for causal evaluation of education policies and practices: a non-technical guide." *CESifo Working Paper 2877*.

SCHMID, E., und STALDER, B. E., (2008). *Lehrvertragsauflösung: Chancen und Risiken für den weiteren Ausbildungsweg*. Erziehungsdirektion des Kantons Bern, Bern.

SCHMID, M., und STORNI, M., (2004). *Transition. Zur Qualifikation und Rekrutierung von Lehrlingen im Kanton Basel-Stadt*. ecce Gemeinschaft für Sozialforschung, Basel.

SCHULER, H., FUNKE, U., und BARON-BOLDT, J., (1990). "Predictive validity of school grades. A meta-analysis." *Applied Psychology*, 39(1), 89-103.

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

SKBF, (2010). *Bildungsbericht 2010*, Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.

SMITH, J., und NAYLOR, R., (2005). "Schooling effects on subsequent university performance: evidence for the UK university population." *Economics of Education Review*, 24(5), 549-562.

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

SPENCE, M., (2002). "Signaling in retrospect and the informational structure of markets." *The American Economic Review*, 92(3), 434-459.

STALDER, B. E., (2000). *Gesucht wird... Rekrutierung und Selektion von Lehrlingen im Kanton Bern*. Amt für Berufsbildung, Bern.

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

STALDER, B. E., und STRICKER, C., (2009). "Traditionelle Arbeitstugenden sind nach wie vor zentral." *Panorama*, 24(5), 14 f.

STAMM, M., NIEDERHAUSER, M., und MÜLLER, R., (2009). *Begabung und Leistungsexzellenz in der Berufsbildung. Eine empirische Studie zu den Ausbildungsverläufen besonders befähigter Jugendlicher im Schweizer Berufsbildungssystem*. Berufsbildungsforschung des BBT, Bern.

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.

TRAPMANN, S., HELL, B., WEIGAND, S., und SCHULER, H., (2007). "Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse." *Zeitschrift für pädagogische Psychologie*, 21(1), 11-27.

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.

WOOLDRIDGE, J. M., (2002a). *Econometric analysis of cross section and panel data*, Cambridge: The MIT Press.

WOOLDRIDGE, J. M., (2002b). *Introductory econometrics: a modern approach*, 2. Auflage, Mason: Thomson/South-Western.

WOOLDRIDGE, J. M., (2003). "Cluster-sample methods in applied econometrics." *The American Economic Review*, 93(2), 133-138.

WÖSSMANN, L., (2005). "The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA." *Education Economics*, 13(2), 143-169.

ZIMMERMAN, D. J., (2003). "Peer effects in academic outcomes: evidence from a natural experiment." *The Review of Economics and Statistics*, 85(1), 9-23.

6.2 INTERNETQUELLEN

BERNER KMU, (2010). *Multicheck oder Note? Ein Leitfaden für Lehrbetriebe*, http://www.lebe.ch/lebe/de/paedagogik/positionspapiere/mainColumnParagraphs/04/document/multicheck_oder_note.pdf. Stand: 09.09.2010.

BFS, (2010). *Räumliche Gliederungen*, http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/geo.html, Neuenburg: BFS. Stand: 07.08.2010.

ILO, (2010). ISCO – International standard classification of occupations. Design and structure, http://www.ilo.org/public/english/bureau/stat/isco/isco88/publ3.htm. Stand: 09.08.2010.

MULTICHECK, (2008). *Zum Zusammenhang zwischen den Ergebnissen des Multichecks und des Qualifikationsverfahrens*, http://multicheck.potentials.ch/fileadmin/user_upload/Wissenschaftliche_Dokumente/Zusammenhang_MC-QV_DH.pdf, Konolfingen: Multicheck. Stand: 05.07.2010.

MULTICHECK, (2010a). *Die Auswertung in der Übersicht*, http://multicheck.potentials.ch/uploads/media/Interpretation_Auswertung.pdf, Konolfingen: Multicheck. Stand: 23.08.2010.

MULTICHECK, (2010b). *Multicheck-Musterauswertung Detailhandel und Service*, http://multicheck.potentials.ch/fileadmin/user_upload/Muster/Zertifikat_Jun_DH.pdf, Konolfingen: Multicheck. Stand: 01.08.2010.

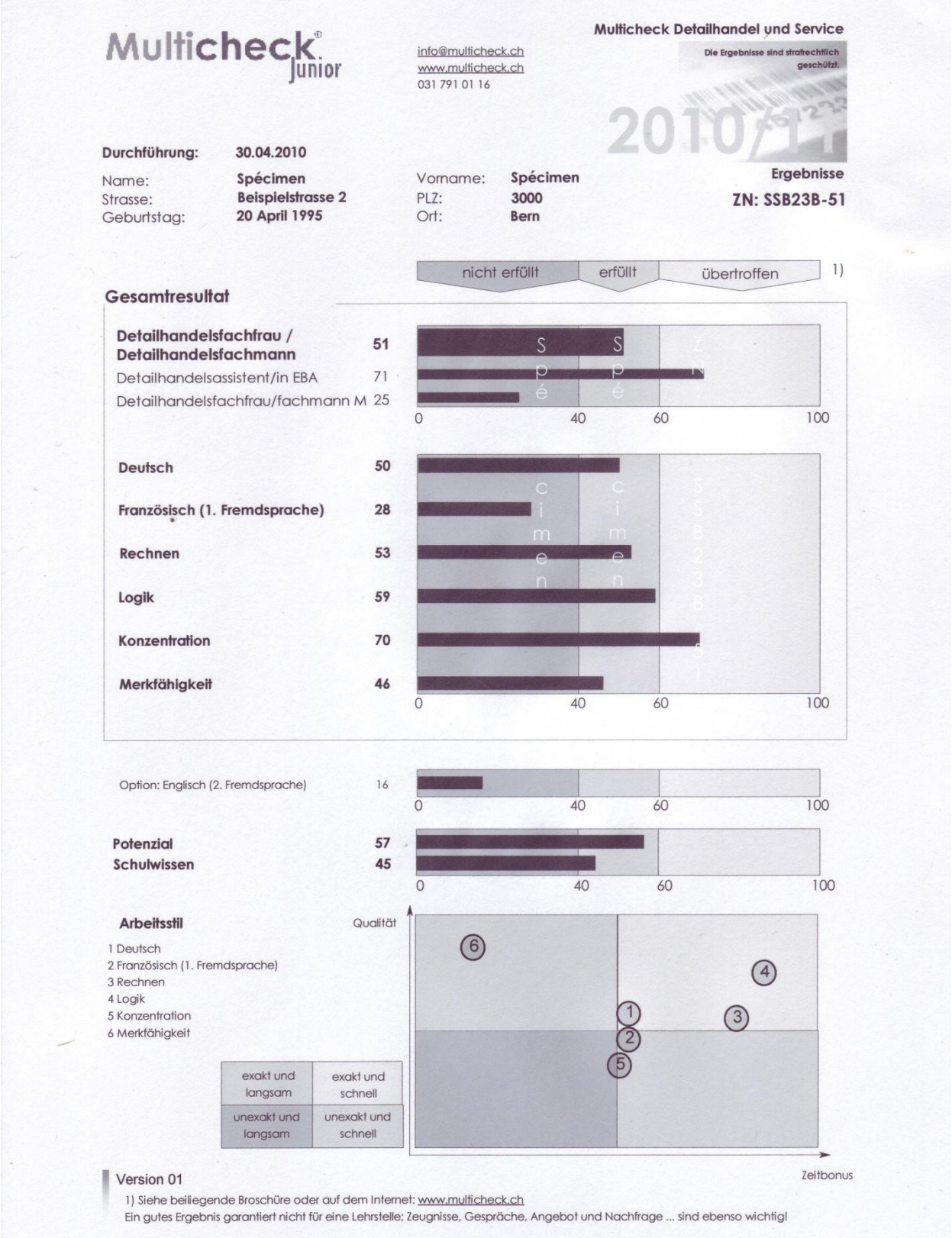
MULTICHECK, (2010c). *Multicheck – Wissenschaftliche Erhebung*, http://multicheck.potentials.ch/fileadmin/user_upload/Wissenschaftliche_Dokumente/Broschuere_Wissenschaft.pdf, Konolfingen: Multicheck. Stand: 23.08.2010.

MULTICHECK, (2010d). *Multicheck Junior*, http://multicheck.potentials.ch/fileadmin/user_upload/Broschueren/broschuere_d.pdf, Konolfingen: Multicheck. Stand: 23.08.2010.

7. ANHANG

A. MULTICHECK-MUSTERAUSWERTUNG

Quelle: Multicheck 2010b



Multicheck® Junior

info@multicheck.ch
www.multicheck.ch
031 791 01 16

Multicheck Detailhandel und Service

Die Ergebnisse sind strafrechtlich geschützt.

2010/11

Ergebnisse

ZN: SSB23B-51

Durchführung:	30.04.2010		
Name:	Spécimen	Vorname:	Spécimen
Strasse:	Beispielstrasse 2	PLZ:	3000
Geburtstag:	20 April 1995	Ort:	Bern

nicht erfüllt | erfüllt | übertroffen 1)

Gesamtresultat

Detailhandelsfachfrau / Detailhandelsfachmann	**51**
Detailhandelsassistent/in EBA	71
Detailhandelsfachfrau/fachmann M	25
Deutsch	**50**
Französisch (1. Fremdsprache)	**28**
Rechnen	**53**
Logik	**59**
Konzentration	**70**
Merkfähigkeit	**46**
Option: Englisch (2. Fremdsprache)	16
Potenzial	**57**
Schulwissen	**45**

Arbeitsstil

1 Deutsch
2 Französisch (1. Fremdsprache)
3 Rechnen
4 Logik
5 Konzentration
6 Merkfähigkeit

Version 01

1) Siehe beiliegende Broschüre oder auf dem Internet: www.multicheck.ch
Ein gutes Ergebnis garantiert nicht für eine Lehrstelle; Zeugnisse, Gespräche, Angebot und Nachfrage ... sind ebenso wichtig!

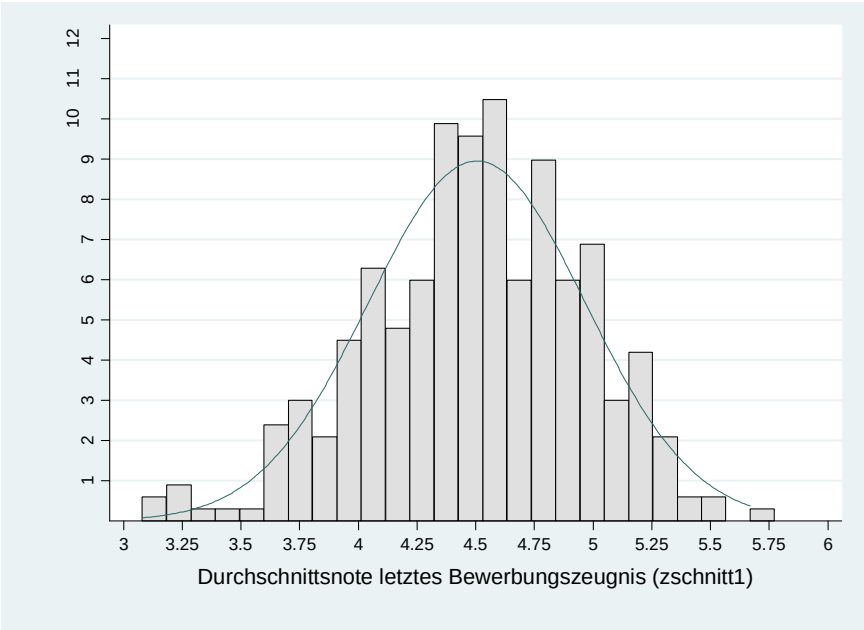
B. VARIABLEN UND CODIERUNG

TABELLE 17: AUS DEN PERSONALDOSSIERS ERHOBENE KENNGRÖSSEN UND DEREN STATA-BEZEICHNUNG (KURSIV)

Persönliche Merkmale (CV)	**Volksschulzeugnisse**	**Multicheck-Auswertung**	**Berufsschulzeugnisse**
- Geschlecht, binär (*weiblich*) - Muttersprache, binär (*muttersp*) - Besuch einer berufsvorbereitenden Schule, binär (*bvs*) - Geburtstag - Beruf des Vaters/der Mutter, nach ÖISCO-Code - Kantonale Herkunft, binär (*ag, be, lu, nw, ow, sg, sz, so, ur, zg, zh,* Diverse Kantone: *divkan)*	- PLZ der Schule - Schulniveau, je binär (*real, sek, gym*) - Schuljahr letztes Bewerbungszeugnis - Noten letztes/neuestes Bewerbungszeugnis (Deutsch: *zd1,* Französisch: *zf1,* Englisch: *ze1,* Math: *zmath1,* Naturwissenschaften: *znat1,* Geschichte und Geo: *zgsgeo1*), Notendurchschnitt (*zschnitt1*) - Absenzen letztes/neuestes Bewerbungszeugnis (*zentsch1, zunentsch1*) - Noten zweitletztes Bewerbungszeugnis (*zd2, zf2, ze2, zmath2, znat2, zgsgeo2*), Notendurchschnitt (*zschnitt2*) - Absenzen zweitletztes Bewerbungszeugnis (*zentsch2, zunentsch2*)	- Durchführungsdatum (*mcdat*) - Testwiederholung, binär (*mcrep*) - Gesamtergebnis (*mcgesamt*) - Einzelne Testgebiete (Deutsch: *mcd*, 1. Fremdsprache: *mcf* oder *mce*, Rechnen: *mcrechnen*, Logik: *mclogik*, Konzentration: *mckonz*, 2. Fremdsprache: *mcf* oder *mce*) - Potenzial (*mcpot*) - Schulwissen (*mcschul*)	- Lehrbeginn - Migros-Genossenschaft, binär (*aare, luzern, zuerich*) - Name der Berufsschule, codiert (*bscode*) - Berufsschulnoten 1. Semester (Deutsch: *ld1*; Fremdsprache: *lf1* oder *le1*, Detailhandelskenntnisse: *ldetail1*, Gesellschaft: *lgesell1*, Wirtschaft: *lwirt1*, Allgemeine Branchenkunde: *lbranch1*), Notendurchschnitt *lschnitt1* - Berufsschulabsenzen 1. Semester (*lentsch1*, *lunentsch1*) - Berufsschulnoten 3. Semester (Deutsch: *ld3*; Fremdsprache: *lf3* oder *le3*, Detailhandelskenntnisse: *ldetail3*, Gesellschaft: *lgesell3*, Wirtschaft: *lwirt3*, Allgemeine Branchenkunde: *lbranch3*), Notendurchschnitt *lschnitt3* - Berufsschulabsenzen 3. Semester (*lentsch3*, *lunentsch3*) - Lehrvertragsauflösung, binär (*labbruch*)

C. HISTOGRAMME EINIGER WICHTIGER VARIABLEN

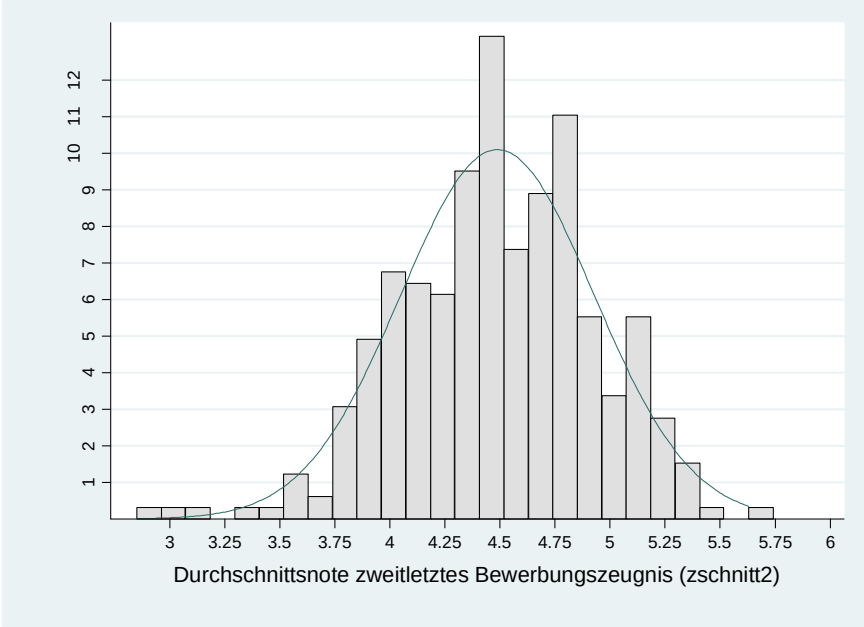
ABBILDUNG 13: HISTOGRAMM DER DURCHSCHNITTSNOTE DES LETZTEN BEWERBUNGSZEUGNISSES



$\mu = 4.502$

$\sigma = 0.461$

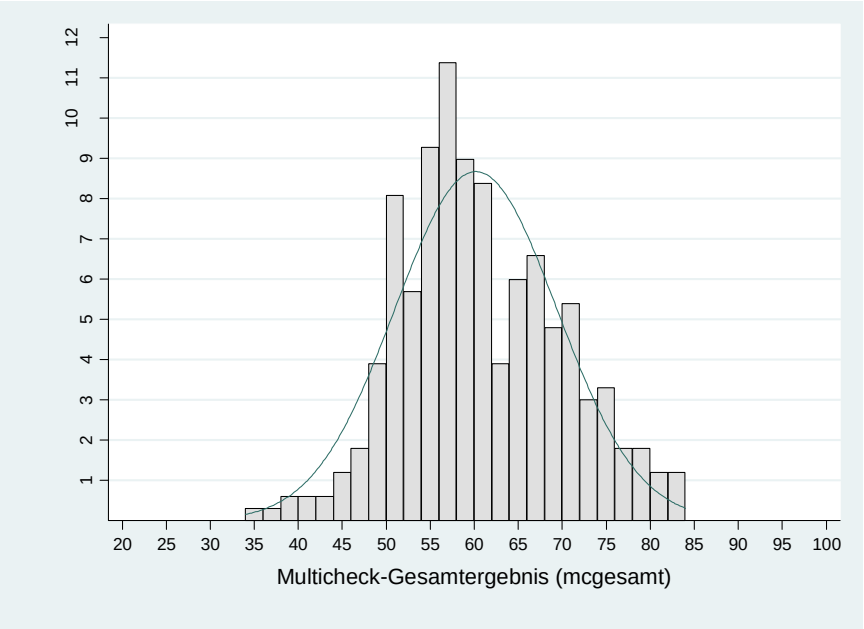
ABBILDUNG 14: HISTOGRAMM DER DURCHSCHNITTSNOTE DES ZWEITLETZTEN BEWERBUNGSZEUGNISSES



$\mu = 4.492$

$\sigma = 0.442$

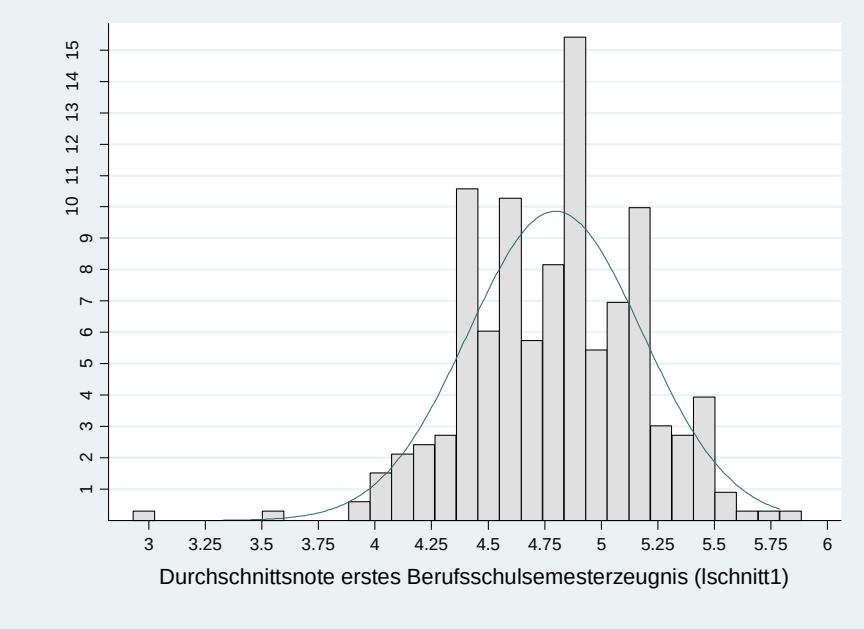
ABBILDUNG 15: HISTOGRAMM DES MULTICHECK-GESAMTERGEBNISSES



μ = 60.17

σ = 9.148

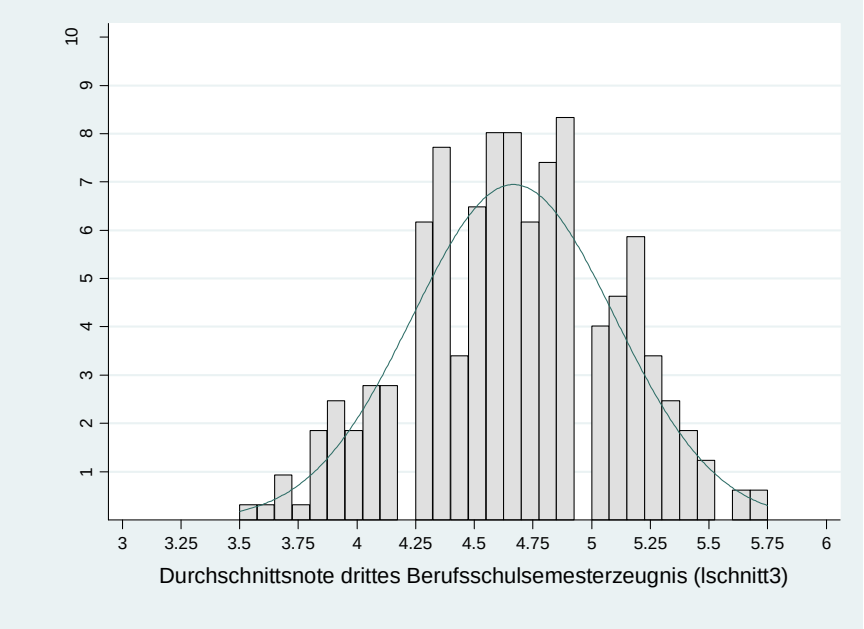
ABBILDUNG 16: HISTOGRAMM DER DURCHSCHNITTLICHEN BERUFS-SCHULNOTEN IM ERSTEN LEHRSEMESTER



μ = 4.806

σ = 0.357

ABBILDUNG 17: HISTOGRAMM DER DURCHSCHNITTLICHEN BERUFSSCHULNOTEN IM DRITTEN LEHRSEMESTER



$\mu = 4.66$

$\sigma = 0.429$

D. OLS- und GLS-Regressionen des Basismodells und des erweiterten Basismodells

TABELLE 18: OLS-REGRESSIONEN DER BERUFSSCHULNOTEN (BASISMODELL UND ERWEITERTES BASISMODELL)

	lschnitt1		*lschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweit rtes Basismodell
mcgesamt	0.00256	0.00342+	0.00237	0.00263
	(0.00208)	(0.00206)	(0.00252)	(0.00254)
	[.002161]	[.002121]	[.002567]	[.002823]
zschnitt12	0.287**	0.289**	0.323**	0.342**
	(0.0435)	(0.0463)	(0.0529)	(0.0570)
	[.039971]	[.031855]	[.04724]	[.046962]
laltera	0.0741	0.0396	-0.0484	-0.0789
	(0.0517)	(0.0498)	(0.0635)	(0.0619)
lalterj	0.0480	0.0312	-0.00106	-0.0199
	(0.0721)	(0.0705)	(0.0874)	(0.0867)
weiblich	0.0233	0.00269	0.0362	0.0235
	(0.0379)	(0.0376)	(0.0462)	(0.0465)
muttersp	0.139**	0.133**	0.0762	0.0719
	(0.0414)	(0.0403)	(0.0509)	(0.0502)
stadt	-0.0646	-0.0738	-0.113*	-0.128*
	(0.0465)	(0.0459)	(0.0568)	(0.0569)
bvs	0.0385	0.0672+	-0.00475	0.0220
	(0.0415)	(0.0406)	(0.0505)	(0.0503)
sek	0.239**	0.283**	0.237**	0.320**
	(0.0399)	(0.0443)	(0.0485)	(0.0548)
gym	0.408**	0.553**	0.519**	0.729**
	(0.121)	(0.122)	(0.147)	(0.150)
Konstante	3.113**	2.836**	2.906**	2.544**
Observationen	329	329	324	324
Korr. R-Quadrat	0.242	0.314	0.195	0.250

Kantons- und Regionsdummies nicht dargestellt (Modell 2 und 4)
Standardfehler in runden Klammern, Cluster-robuste Standardfehler in eckigen Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 19: CLUSTER-ROBUSTE GLS-REGRESSIONEN DER BERUFSSCHULNOTEN (BASISMODELL UND ERWEITERTES BASISMODELL)

	lschnitt1		*lschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweitertes Basismodell
mcgesamt	0.00461*	0.00488**	0.00207	0.00214
	(0.00185)	(0.00171)	(0.00224)	(0.00214)
zschnitt12	0.313**	0.296**	0.343**	0.359**
	(0.0377)	(0.0393)	(0.0477)	(0.0480)
Observationen	329	329	324	324
Cluster	23	23	23	23

Restlicher Regressionsoutput nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

E. Signalwert nach Multicheck-Durchführungsjahr und Schuljahr des Zeugnisses

TABELLE 20: FIXED-EFFECTS-REGRESSIONEN DES BASISMODELLS NACH MULTICHECKVERSION

	2006/2007		2007/2008	
	lschnitt1	*lschnitt3*	*lschnitt1*	*lschnitt3*
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Basismodell	Basismodell	Basismodell
mcgesamt	0.00126	0.00130	0.00465	0.00355
	(0.00300)	(0.00309)	(0.00418)	(0.00739)
zschnitt12	0.260**	0.169*	0.366**	0.525**
	(0.0462)	(0.0808)	(0.0719)	(0.115)
Observationen	184	183	109	109
R-Quadrat	0.249	0.215	0.332	0.288
Cluster	23	23	20	20

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 21: PRÄDIKTIVE VALIDITÄT EINZELNER SEMESTERZEUGNISSE IN DEN FIXED-EFFECTS-REGRESSIONEN DES ERWEITERTEN BASISMODELLS

	lschnitt1				*lschnitt3*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLEN								
mcgesamt	0.00193	0.00267	0.00708**	0.00701*	-0.00422	0.000389	0.00625+	0.00601
	(0.00396)	(0.00178)	(0.00198)	(0.00249)	(0.00522)	(0.00281)	(0.00343)	(0.00446)
zschnitt95	0.165**				0.304**			
	(0.0572)				(0.0743)			
zschnitt9		0.153**				0.206*		
		(0.0371)				(0.0760)		
zschnitt85			0.267**				0.330**	
			(0.0539)				(0.0885)	
zschnitt8				0.334**				0.356**
				(0.0747)				(0.0641)
Observationen	133	164	174	138	131	160	173	137
R-Quadrat	0.300	0.217	0.301	0.334	0.233	0.187	0.304	0.293
Cluster	20	20	23	23	20	20	23	23

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

F. NORMALQUANTIL-PLOTS DER RESIDUEN DER FIXED-EFFECTS-REGRESSIONEN

ABBILDUNG 18: NORMALQUANTIL-PLOT DER RESIDUEN DES FIXED-EFFECTS-MODELLS FÜR LSCHNITT1 (ERWEITERTES BASISMODELL)

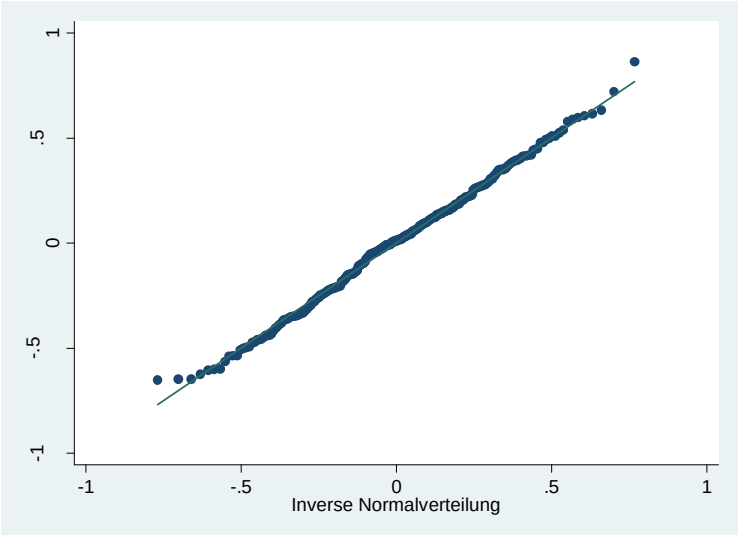
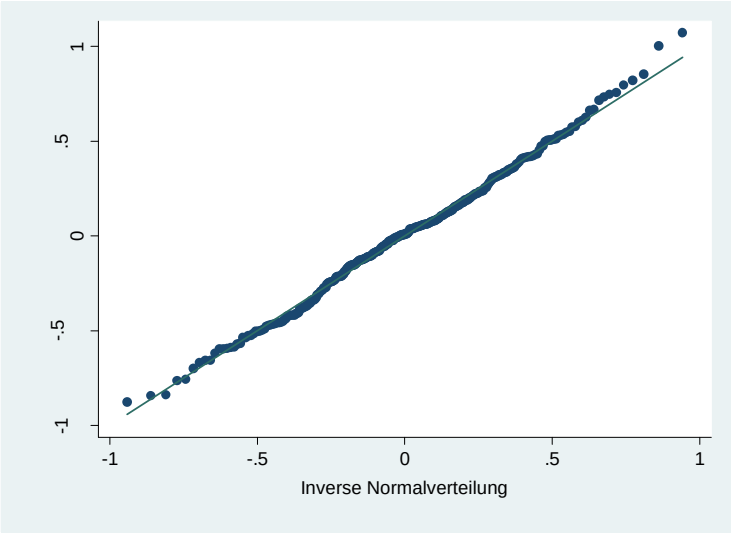


ABBILDUNG 19: NORMALQUANTIL-PLOT DER RESIDUEN DES FIXED-EFFECTS-MODELLS FÜR LSCHNITT3 (ERWEITERTES BASISMODELL)



G. Ordered-Probit-Regressionen unter Kontrolle der sozialen Herkunft der Lernenden

TABELLE 22: ORDERED-PROBIT-MODELLE DER BERUFSSCHULNOTEN DES ERSTEN UND DRITTEN SEMESTERS UNTER VERWENDUNG DER SOZIALEN HERKUNFT DER LERNENDEN

	lquartillschnitt1		*lquartillschnitt3*	
VARIABLEN	(1) Erweitertes Basismodell	(2) Erweitertes Basismodell inkl. *berufe12*	(3) Erweitertes Basismodell	(4) Erweitertes Basismodell inkl. *berufe12*
mcgesamt	0.0158+	0.0151	0.0133	0.0144
	(0.00913)	(0.00924)	(0.00910)	(0.00924)
zschnitt12	0.938**	0.937**	0.722**	0.724**
	(0.211)	(0.211)	(0.210)	(0.209)
berufe12		0.0802		-0.111
		(0.164)		(0.164)
laltera	-0.0973	-0.109	-0.384+	-0.367
	(0.228)	(0.230)	(0.231)	(0.233)
lalterj	-0.119	-0.122	0.0192	0.0221
	(0.285)	(0.285)	(0.287)	(0.287)
weiblich	-0.0509	-0.0404	0.0927	0.0784
	(0.167)	(0.169)	(0.169)	(0.170)
muttersp	0.313+	0.291	-0.00205	0.0264
	(0.178)	(0.183)	(0.180)	(0.185)
stadt	-0.312	-0.307	-0.273	-0.282
	(0.213)	(0.213)	(0.212)	(0.213)
bvs	0.0275	0.0251	-0.0201	-0.0176
	(0.176)	(0.176)	(0.177)	(0.177)
sek	0.862**	0.864**	0.728**	0.727**
	(0.202)	(0.202)	(0.201)	(0.202)
gym	1.807**	1.785**	2.454**	2.480**
	(0.543)	(0.545)	(0.609)	(0.607)
luzern	0.110	0.0878	1.235	1.274
	(0.962)	(0.961)	(0.967)	(0.972)
zuerich	-0.214	-0.227	0.353	0.372
	(0.522)	(0.522)	(0.512)	(0.514)
α_1	5.272**	5.243**	3.765**	3.814**
α_2	6.974**	6.947**	5.460**	5.510**
Observationen	247	247	243	243
Pseudo-R-Quadrat	0.112	0.113	0.0995	0.100

Kantonsdummies nicht dargestellt (Modelle 2 und 4)
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

H. AUFBAUENDE ORDERED-PROBIT-MODELLE

TABELLE 23: SIGNIFIKANZ UND EFFEKTGRÖSSE DES MULTICHECK-GESAMTERGEBNISSES IM ORDERED-PROBIT-MODELL UNTER VERWENDUNG EINES IMMER GRÖSSEREN SATZES AN KONTROLLVARIABLEN

	lquartillschnitt1				*lquartillschnitt3*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLEN								
mcgesamt	0.0175*	0.00906	0.0101	0.00887	0.0199**	0.0130+	0.0140+	0.0133+
	(0.0100)	(0.205)	(0.168)	(0.248)	(0.00379)	(0.0711)	(0.0588)	(0.0864)
bvs		0.0877	0.0573	0.0743		-0.0153	0.0412	0.0458
		(0.507)	(0.696)	(0.624)		(0.909)	(0.779)	(0.764)
sek		0.499**	0.432**	0.624**		0.359**	0.319*	0.504**
		(0.000199)	(0.00188)	(9.14e-5)		(0.00757)	(0.0225)	(0.00161)
gym		1.028*	0.932*	1.340**		1.272**	1.321**	1.651**
		(0.0181)	(0.0342)	(0.00427)		(0.00516)	(0.00467)	(0.000731)
laltera			0.114	0.0427			-0.170	-0.220
			(0.532)	(0.818)			(0.360)	(0.242)
lalterj			-0.0425	-0.0391			-0.0220	0.0253
			(0.868)	(0.882)			(0.932)	(0.924)
weiblich			0.0946	0.0767			0.169	0.187
			(0.471)	(0.580)			(0.202)	(0.182)
muttersp			0.315*	0.355*			0.175	0.186
			(0.0317)	(0.0190)			(0.241)	(0.225)
stadt			-0.0636	-0.169			-0.0971	-0.149
			(0.698)	(0.323)			(0.557)	(0.387)
luzern				0.0570				0.198
				(0.899)				(0.443)
zuerich				-0.119				0.230
				(0.763)				(0.396)
α_1	0.305	0.0491	0.355	0.763	0.435	0.184	0.426	0.662
α_2	1.755**	1.547**	1.870**	2.325**	1.991**	1.780**	2.036**	2.317**
Observationen	329	329	329	329	324	324	324	324
Pseudo-R-Quadrat	0.00992	0.0351	0.0439	0.0670	0.0131	0.0341	0.0416	0.0621

Kantonsdummies nicht dargestellt (Modelle 4 und 8)
p-Werte in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 24: SIGNIFIKANZ UND EFFEKTGRÖSSE DER SCHULZEUGNISNOTEN IM ORDERED-PROBIT-MODELL UNTER VERWENDUNG EINES IMMER GRÖSSEREN SATZES AN KONTROLLVARIABLEN

	lquartillschnitt1				*lquartillschnitt3*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLEN								
zschnitt12	0.425**	0.642**	0.650**	1.001**	0.333*	0.525**	0.495**	0.753**
	(0.00348)	(2.77e-5)	(3.90e-5)	(3.83e-8)	(0.0224)	(0.00065)	(0.0017)	(2.43e-5)
bvs		0.165	0.120	0.137		0.0552	0.101	0.1000
		(0.215)	(0.412)	(0.370)		(0.680)	(0.492)	(0.515)
sek		0.663**	0.603**	0.995**		0.514**	0.473**	0.811**
		(7.60e-7)	(1.37e-5)	(2.05e-9)		(0.00012)	(0.00065)	(7.60e-7)
gym		1.461**	1.338**	2.082**		1.704**	1.710**	2.275**
		(0.0009)	(0.00259)	(2.10e-5)		(0.00025)	(0.00032)	(6.15e-6)
laltera			0.153	0.118			-0.155	-0.178
			(0.403)	(0.530)			(0.404)	(0.348)
lalterj			-0.0690	-0.0886			-0.0297	0.00451
			(0.788)	(0.739)			(0.908)	(0.986)
weiblich			-0.0525	-0.0740			0.0385	0.0600
			(0.692)	(0.599)			(0.772)	(0.671)
muttersp			0.304*	0.393*			0.177	0.220
			(0.0389)	(0.0102)			(0.236)	(0.151)
stadt			-0.0124	-0.191			-0.0516	-0.165
			(0.940)	(0.270)			(0.756)	(0.344)
luzern				0.0250				0.146
				(0.956)				(0.744)
zuerich				0.0744				0.385
				(0.854)				(0.340)
α_1	1.163+	2.480**	2.671**	4.882**	0.750	1.862**	1.839*	3.384**
α_2	2.618**	4.027**	4.236**	6.548**	2.295**	3.484**	3.468**	5.086**
Observationen	329	329	329	329	324	324	324	324
Pseudo-R-Quadrat	0.0128	0.0593	0.0667	0.112	0.0081	0.0474	0.0515	0.0856

Kantonsdummies nicht dargestellt (Modelle 4 und 8)
p-Werte in Klammern
** p<0.01, * p<0.05, + p<0.1

I. SCHULNOTENSIGNALWERT NACH SCHULTYP

TABELLE 25: FIXED-EFFECTS-REGRESSIONEN MIT INTERAKTIONSTERMEN ZWISCHEN DEN DURCHSCHNITTLICHEN SCHULNOTEN DER BEWERBUNGSZEUGNISSE UND DEM SCHULTYP

	lschnitt1		*lschnitt3*	
	(1)	(2)	(3)	(4)
VARIABLEN	Basismodell	Erweitertes Basismodell	Basismodell	Erweitertes Basismodell
zschnitt12	0.283**	0.289**	0.337**	0.343**
	(0.0283)	(0.0255)	(0.0715)	(0.0722)
zschnitt12sek	0.0853	0.0642	0.0507	0.0594
	(0.0602)	(0.0637)	(0.130)	(0.135)
Observationen	317	317	313	313
R-Quadrat	0.267	0.294	0.209	0.247
Cluster	23	23	23	23

Restlicher Regressionsoutput nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

J. FIXED-EFFECTS-MODELLE DER NOTEN EINZELNER BERUFSSCHULFÄCHER

TABELLE 26: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH DEUTSCH

	ld13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.00711*	
	(0.00285)	
zschnitt12	0.230**	
	(0.0614)	
mcd		0.00380**
		(0.00122)
zd12		0.234*
		(0.0930)
laltera	-0.0935	-0.0908+
	(0.0574)	(0.0500)
lalterj	0.0690	0.0398
	(0.0673)	(0.0638)
weiblich	0.126**	0.116**
	(0.0233)	(0.0409)
muttersp	0.153**	0.0964*
	(0.0461)	(0.0397)
stadt	-0.0573	-0.0663
	(0.0506)	(0.0489)
bvs	-0.0270	-0.0220
	(0.0387)	(0.0407)
sek	0.278**	0.296**
	(0.0392)	(0.0382)
gym	0.541**	0.522**
	(0.154)	(0.146)
Konstante	2.709**	2.824**
Observationen	330	330
R-Quadrat	0.253	0.270
AIC	285.9	278.4
BIC	361.9	354.4
Cluster	23	23

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 27: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH FRANZÖSISCH

	lf13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.00384	
	(0.00699)	
zschnitt12	0.288+	
	(0.160)	
mcf		0.0144**
		(0.00436)
zf12		0.257*
		(0.107)
laltera	-0.182	-0.0891
	(0.141)	(0.168)
lalterj	0.367+	0.165
	(0.206)	(0.167)
weiblich	0.0919	0.0890
	(0.122)	(0.144)
muttersp	-0.0600	-0.159
	(0.0944)	(0.113)
stadt	0.0771	-0.0645
	(0.163)	(0.178)
bvs	-0.0596	-0.189
	(0.166)	(0.130)
sek	0.283	0.0963
	(0.173)	(0.178)
gym	0.804**	0.479*
	(0.207)	(0.213)
Konstante	2.585**	2.400**
Observationen	117	117
R-Quadrat	0.200	0.347
AIC	192.7	169.0
BIC	223.1	199.4
Cluster	17	17

Kantons- und Regionsdummies nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 28: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH ENGLISCH

	le13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.0100+	
	(0.00483)	
zschnitt12	0.393**	
	(0.0689)	
mce		0.00771**
		(0.00128)
ze12		0.293**
		(0.0414)
laltera	0.126	0.193+
	(0.0931)	(0.103)
lalterj	-0.105	-0.117
	(0.139)	(0.111)
weiblich	0.124+	0.0906*
	(0.0627)	(0.0355)
muttersp	-0.117	-0.129
	(0.0927)	(0.0747)
stadt	-0.196**	-0.187*
	(0.0664)	(0.0685)
bvs	0.180	0.119
	(0.143)	(0.129)
sek	0.538**	0.535**
	(0.0512)	(0.0898)
gym	0.634**	0.515**
	(0.0399)	(0.0787)
Konstante	1.899**	2.564**
Observationen	174	174
R-Quadrat	0.332	0.403
AIC	209.3	189.8
BIC	250.3	230.9
Cluster	19	19

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 29: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH DETAILHANDELSKENNTNISSE

	ldetail13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.000306	
	(0.00260)	
zschnitt12	0.313**	
	(0.0575)	
mcmerk		0.00280*
		(0.00121)
zgsgeo12		0.163**
		(0.0343)
zmath12		0.101**
		(0.0328)
laltera	-0.0256	-0.0508
	(0.0852)	(0.0842)
lalterj	-0.0105	-0.00752
	(0.109)	(0.0976)
weiblich	-0.0556	-0.0247
	(0.0450)	(0.0416)
muttersp	0.165**	0.175**
	(0.0567)	(0.0532)
stadt	-0.0559	-0.0435
	(0.0606)	(0.0614)
bvs	0.0281	0.0381
	(0.0614)	(0.0589)
sek	0.209**	0.191**
	(0.0473)	(0.0426)
gym	0.583**	0.581**
	(0.143)	(0.144)
Konstante	3.018**	3.094**
Observationen	327	327
R-Quadrat	0.164	0.187
AIC	339.6	332.4
BIC	415.4	412.0
Cluster	23	23

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 30: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN BERUFSSCHULNOTE IM FACH BRANCHENKUNDE

	lbranch1	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.00481	
	(0.00380)	
zschnitt12	0.444**	
	(0.0610)	
mcrechnen		0.00480*
		(0.00178)
zf12		0.115*
		(0.0540)
znat12		0.248**
		(0.0733)
laltera	-0.116	-0.136
	(0.0787)	(0.102)
lalterj	0.00145	0.0569
	(0.124)	(0.117)
weiblich	0.0105	0.0510
	(0.0819)	(0.0793)
muttersp	0.150	0.156
	(0.104)	(0.113)
stadt	-0.0704	-0.0659
	(0.0918)	(0.0868)
bvs	0.0781	0.0806
	(0.0943)	(0.103)
sek	0.310**	0.255*
	(0.0972)	(0.101)
gym	0.766**	0.707**
	(0.224)	(0.209)
Konstante	1.804**	2.207**
Observationen	295	295
R-Quadrat	0.219	0.242
AIC	446.8	440.2
BIC	520.5	517.7
Cluster	22	22

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 31: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH GESELLSCHAFT

	lgesell13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.000513	
	(0.00296)	
zschnitt12	0.303**	
	(0.0605)	
mce		0.00410**
		(0.00141)
znat12		0.289**
		(0.0640)
zd12		-0.145*
		(0.0570)
laltera	-0.0242	-0.0756
	(0.0806)	(0.0918)
lalterj	0.0321	-0.133
	(0.119)	(0.0810)
weiblich	-0.0587	0.0173
	(0.0491)	(0.0383)
muttersp	0.0964	0.117
	(0.0631)	(0.0691)
stadt	-0.149+	-0.188*
	(0.0851)	(0.0735)
bvs	0.162*	0.127
	(0.0775)	(0.0754)
sek	0.207*	0.158*
	(0.0775)	(0.0748)
gym	0.394+	0.300
	(0.216)	(0.205)
Konstante	3.163**	3.700**
Observationen	330	292
R-Quadrat	0.166	0.253
AIC	403.3	321.2
BIC	479.3	391.0
Cluster	23	23

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

TABELLE 32: FIXED-EFFECTS-MODELLE DER DURCHSCHNITTLICHEN ERST- UND DRITTSEMESTER-BERUFSSCHULNOTE IM FACH WIRTSCHAFT

	lwirt13	
	(1)	(2)
VARIABLEN	Erweitertes Basismodell	Erweitertes Basismodell
mcgesamt	0.00288	
	(0.00325)	
zschnitt12	0.403**	
	(0.0825)	
mce		0.00482**
		(0.00129)
zmath12		0.120**
		(0.0378)
zgsgeo12		0.249**
		(0.0755)
laltera	0.0334	-0.0268
	(0.103)	(0.100)
lalterj	0.0932	0.0187
	(0.139)	(0.144)
weiblich	-0.0407	0.0460
	(0.0728)	(0.0770)
muttersp	0.101	0.106
	(0.0889)	(0.0842)
stadt	-0.163	-0.173+
	(0.102)	(0.0963)
bvs	0.0329	0.0344
	(0.0917)	(0.0827)
sek	0.482**	0.471**
	(0.0917)	(0.0852)
gym	0.751**	0.724**
	(0.227)	(0.189)
Konstante	2.123**	1.773**
Observationen	330	293
R-Quadrat	0.226	0.283
AIC	489.0	411.7
BIC	565.0	481.7
Cluster	23	23

Kantons- und Regionsdummies nicht dargestellt
Cluster-robuste Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

K. RE- UND FE-REGRESSION DER WAHRSCHEINLICHKEIT UNENTSCHULDIGTER ABSENZEN IN DER BERUFSSCHULE

TABELLE 33: FIXED-EFFECTS-, RANDOM-EFFECTS-LOGIT- UND -PROBIT-MODELLE DER UNENTSCHULDIGTEN ABSENZEN

	lunentsch01				
VARIABLEN	(1) Logit	(2) Fixed-Effects-Logit	(3) Random-Effects-Logit	(4) Probit	(5) Random-Effects-Probit
mcgesamt	0.0214	0.00215	0.0140	0.0133	0.00924
	(0.0193)	(0.0229)	(0.0222)	(0.0110)	(0.0127)
zschnitt12	-0.955**	-1.009*	-0.930*	-0.571**	-0.559*
	(0.318)	(0.452)	(0.441)	(0.168)	(0.252)
laltera	0.259	0.403	0.343	0.122	0.178
	(0.362)	(0.487)	(0.462)	(0.204)	(0.265)
lalterj	0.180	0.173	0.180	0.0865	0.0872
	(0.586)	(0.828)	(0.750)	(0.323)	(0.424)
weiblich	-0.473	-0.558	-0.506	-0.261	-0.269
	(0.351)	(0.375)	(0.362)	(0.191)	(0.203)
muttersp	0.0191	0.111	0.0142	-0.0107	-0.0133
	(0.457)	(0.419)	(0.408)	(0.258)	(0.228)
stadt	0.434*	0.628	0.465	0.254*	0.265
	(0.219)	(0.423)	(0.411)	(0.120)	(0.236)
bvs	0.0310	-0.0514	-0.00834	0.0276	-0.00101
	(0.419)	(0.428)	(0.405)	(0.228)	(0.230)
sek	-0.708+	-1.062*	-0.734	-0.400*	-0.415
	(0.379)	(0.494)	(0.446)	(0.201)	(0.253)
gym	0.344	0.163	0.359	0.231	0.231
	(1.231)	(1.120)	(1.059)	(0.746)	(0.633)
Konstante	1.964		2.229	1.137	1.294
Observationen	259	259	259	259	259
Cluster		16	16		16

Kantonsdummies nicht dargestellt
Standardfehler in Klammern
** p<0.01, * p<0.05, + p<0.1

L. SELBSTÄNDIGKEITSERKLÄRUNG

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, den Michael Siegenthaler