

The final publication is available at Hogrefe via DOI:

<http://dx.doi.org/10.1024/1010-0652.19.4.219>

Zeitschrift für Pädagogische Psychologie, 19, 219-224 © 2005
by Dickhäuser, O & Plenter, I. This version of the article may
not completely replicate the final version published in
Zeitschrift für Pädagogische Psychologie. It is not the ver-
sion of record and is therefore not suitable for citation.

Kolummentitel: AKKURATHEIT SELBST BERICHTETER NOTEN

"Letztes Halbjahr stand ich zwei"

Zur Akkuratheit selbst berichteter Noten

Oliver Dickhäuser¹ & Insa Plenter²

¹Friedrich-Alexander-Universität Erlangen-Nürnberg

²Justus-Liebig Universität Giessen

Autorenhinweis

Prof. Dr. Oliver Dickhäuser, Lehrstuhl Pädagogische Psychologie, Erziehungswissenschaftliche Fakultät, Friedrich-Alexander-Universität, Regensburger Straße 160, D-90478 Nürnberg. e-mail: oliver.dickhaeuser@ewf.uni-erlangen.de

Zusammenfassung

Pädagogisch-psychologische Studien verwenden oft die von Schülern selbst berichteten Noten als Leistungsindikatoren. In der vorliegenden Studie wurde die Akkuratheit dieser Selbstangaben überprüft, indem selbst berichtete Mathematiknoten der letzten Klassenarbeit und des letzten Zeugnisses mit den Lehrerangaben verglichen wurden. Es zeigen sich hohe Korrelationen, allerdings wurden die Noten im Mittel überschätzt. Die Akkuratheit der Selbsteinschätzung ist praktisch unabhängig von Geschlecht, mathematischem Fähigkeitsselbstkonzept und mathematischer Leistung (erfasst über Items aus dem TIMSS-Instrumentarium).

Schlagwörter: Leistung - Selbstbeurteilung - Selbstkonzept - Schüler der Sekundarstufe

On the accuracy of self-reported school marks

Summary

Studies in Educational Psychology often use students' self-reports on their school marks as an indicator of students' performance. In the present study we compared self-reports on marks from the last class test and from the last report in mathematics with the marks as reported by the teacher. There was a strong correlation between self-reported and actual marks, however, students overestimated their marks. The accuracy of the self-report did not depend meaningfully on student's gender, mathematical self-concept and mathematical achievement (as measured by a standardized math test).

Keywords: academic-achievement - self-evaluation - academic-self-concept - high school students

1 Theoretischer Hintergrund

In pädagogisch-psychologischen Studien werden oft die von Schülern berichteten Noten als Maß für die Schülerleistung verwendet. Dies ist ein ökonomisches Vorgehen; die Erhebung der Lehrerangaben über die Zensuren ist häufig nur schwer zu realisieren. Zweifelhaft ist aber, ob die Selbstberichte die tatsächlichen Zensuren akkurat widerspiegeln und als valides Leistungsmaß angesehen werden können.

Einige Untersuchungen im angloamerikanischen Raum haben sich mit der Übereinstimmung selbst berichteter und tatsächlicher Leistungsmaße beschäftigt. In einer frühen Untersuchung fanden Kirk und Sereda (1969) bei Studenten für den Notendurchschnitt (grade point average, GPA) Korrelationen von .93 bis 1.00. Bei Goldman, Flake und Matheson (1990) lagen die Korrelationen zwischen .70 und .88. Frucot und Cook (1994) erfassten neben GPA-scores auch selbst berichtete und tatsächliche Einzelnoten. Letztere korrelierten zwischen .90 bis .95. Allerdings wurden die eigenen Leistungen im Mittel eher überschätzt (Frucot & Cook, 1994; Kirk & Sereda, 1969; Goldman, Flake & Matheson, 1990). Uneindeutig ist die Befundlage zu der Frage, ob die Akkuratheit des Selbstberichts von der Leistungshöhe abhängig ist. So zeigten in einer Studie von Dobbins, Farh und Werbel (1993) leistungsschwache College Studenten eine höhere Tendenz zur Überschätzung als leistungsstarke (vgl. auch Shepperd, 1993), während Goldman et al. (1990) in Abhängigkeit der Leistungsstärke keine unterschiedliche Akkuratheit zeigen konnten.

Einen niedrigeren Zusammenhang zwischen selbst berichtetem und tatsächlichem GPA fanden Zimmerman, Caldwell und Bernat (2002) mit $\underline{r} = .62$ in einer Stichprobe von 679 afro-

amerikanischen Schülern. Zwei Besonderheiten sind bei der Interpretation dieses Ergebnisses von Bedeutung. Erstens wurden die Versuchspersonen teils von weißen Versuchsleitern befragt, obwohl die Autoren betonen: "...previous research has indicated that African Americans respond differently with African American interviewers than they do with white interviewers on sensitive issues" (S. 87). Zweitens wurden die Schüler bei der Erfassung des GPA nicht befragt, welche Durchschnittsnote sie im letzten Schuljahr hatten, sondern die Frage lautete "Which grade best describes your average grade last year?" (S. 93, Hervorhebung durch uns). Auf diese Weise wird möglicherweise nicht die erinnerte Durchschnittsnote erfasst, sondern die Note, die dem Leistungsniveau am ehesten als gerecht empfunden wird. Beide Punkte zusammen genommen können möglicherweise die außergewöhnlich niedrige Korrelation erklären.

Diese Befunde aus dem angloamerikanischen Raum können ohnehin nur bedingt auf das deutsche Schulsystem übertragen werden. Es gibt kein dem GPA gleich bedeutsames Maß im Sinne einer Durchschnittsnote: Zeugnisse in den einzelnen Bundesländern enthalten diese Angabe nicht. Dagegen kommt den Noten in einzelnen Fächern wegen der Versetzungsrelevanz nicht ausreichender Leistungen eine ungleich größere Bedeutung zu.

Unseres Wissens wurde der Frage nach der Akkuratheit selbst berichteter Noten im deutschen Schulsystem bislang in drei (Teil-)Studien empirisch nachgegangen. Schelma (1993) befragte Schüler/-innen der Stufen 12 und 13 nach den Zeugnisnoten in Klasse 6 und 9. Diese Erinnerungsspanne ist sehr lang und für die meisten Studien, in denen Schüler Angaben über eigene Leistungen machen untypisch: Meistens werden Schüler darum gebeten, ihre letzte Zeugnisnote sowie die letzte Note der

Klassenarbeit anzugeben. Die Befunde von Schelma liefern auch deshalb kein klares Bild über die Güte der Notenerinnerung, weil diese unter einer von drei Experimentalbedingungen erfasst wurde, die Notenverzerrungen begünstigen oder verhindern sollte (den Versuchspersonen wurde die Begabungs- oder Anstrengungsabhängigkeit schulischer Leistungen nahegelegt oder sie wurden aufgefordert, vor dem Notieren der einstigen Note mögliche Schlüsselereignisse im Zusammenhang mit dieser zu erinnern). Helmke (1992) fand bei Hauptschülern klassenweise Korrelationen zwischen selbst berichteten und vom Lehrer angegebenen Zeugnisnoten im Bereich von .75 und .97. In einer Zusatzstudie der Arbeit von Möller, Streblow, Pohlmann und Köllner (2003) wurden 83 Gymnasiasten gebeten, die Mathematiknote ihrer letzten Klassenarbeit zu erinnern. Der Zusammenhang zwischen selbst berichteter und vom Lehrer angegebener Note betrug .93. Zehn Schüler überschätzten, sechs unterschätzten ihre Note. Die Güte des Selbstberichts war unabhängig vom mathematischen Fähigkeitsselbstkonzept.

In der vorliegenden Studie soll in einer großen Stichprobe überprüft werden, inwieweit Schülerangaben zu Noten im letzten Zeugnis und in der letzten Klassenarbeit mit den Noten der Lehrkräfte korrespondieren. Anders als in der Arbeit von Möller et al. (2003) wird nicht nur die Note aus der letzten Klassenarbeit untersucht, sondern auch die des letzten Zeugnisses. Wir wählen eine Erinnerungsspanne, die der gängigen Praxis in Schuluntersuchungen eher entspricht als das bei Schelma (1993) gewählte Intervall. Im einzelnen werden folgende Fragen überprüft:

- Wie eng hängen selbst berichtete Note und Lehrerangabe der Note zusammen?

- Entspricht das Niveau der selbst berichteten Note dem der Lehrerangabe und werden etwaige Abweichungen moderiert durch Fähigkeitsselbstkonzept, Geschlecht oder Leistungsniveau?
- Sind selbst berichtete und vom Lehrer angegebene Note gleich valide Leistungsmaße?

2.1 Methode

Die Stichprobe bestand aus 866 Schülern (davon 414 weiblich) eines Gymnasiums ($n = 241$), dreier Realschulen ($n = 397$) sowie zwei Hauptschulen ($n = 228$) aus Nordrhein-Westfalen. Die Schüler verteilten sich auf 36 verschiedene Klassen, wobei insgesamt 494 der 7. Klassenstufe, 372 der 8. Klassenstufe angehörten. Das mittlere Alter betrug 13.75 Jahre ($SD = 0.85$).

Die Durchführung der Befragung erfolgte während regulärer Schulstunden durch geschulte Versuchsleiter und dauerte 45 bis 70 Minuten. Als Material zur Erfassung des mathematischen Leistungsniveaus bearbeiteten die Schüler eine Auswahl von 28 Items der Testaufgaben Mathematik der Third International Mathematics and Science Study für 7./8. Klassen (Population 2; Baumert et al., 1998). Die Itemauswahl erwies sich in einer Voruntersuchung als zufriedenstellend messgenau und valide. In der vorliegenden Stichprobe betrug die Split-Half-Reliabilität $r = .77$. Die Siebtklässler lösten 16.1, Achtklässler 17.4 der 28 Aufgaben korrekt. Nachfolgend notierten die Schüler demographische Angaben sowie die Noten der letzten Klassenarbeit in Mathematik (einschließlich der Notentendenz) und des letzten Zeugnisses. Die Klassenarbeit lag etwa drei Wochen, das Zeugnis etwa viereinhalb Monate zurück. Unter anderem wurden dann die Items der Skala "absolutes mathematisches Fähigkeitsselbstkonzept" vorgegeben (vgl. Dickhäuser, 2003; Beispieli-tem: "Ich bin für Mathe ..." "nicht begabt" [1] vs. "sehr be-

gabt" [5]). Die interne Konsistenz betrug .92. Die Noten wurden auch per Lehrerangabe erfasst und mittels anonymer Personencodes zugeordnet.

2.2 Ergebnisse

Die Mittelwerte (und Standardabweichungen) der selbst berichteten und vom Lehrer berichteten Noten fielen wie folgt aus: Selbstbericht Zeugnis 3.34 (0.98) bzw. Klassenarbeit 3.42 (1.22), Lehrerangabe Zeugnis 3.43 (1.00) bzw. Klassenarbeit 3.51 (1.22). Die Korrelation zwischen selbst berichteter Note und Lehrerangabe betrug für die Klassenarbeit $r = .90$ und für die Zeugnisnote $r = .88$ (Differenz n.s.). Das Niveau der Akkuratheit wurde mittels zweier t-Tests für gepaarte Stichproben geprüft, bei denen Selbstangabe und Lehrerangabe miteinander verglichen wurden. Die Noten fielen für Klassenarbeit (t [835] = 5.52, $p < .001$, $ES = .09$) und Zeugnis (t [844] = 4.98, $p < .001$, $ES = .09$) im Selbstbericht besser aus als die Lehrerangabe. Die Überschätzung beträgt im Mittel bei der Klassenarbeit 0.10 und bei der Zeugnisnote 0.09 Notenstufen. Bei der Zeugnisnote [Klassenarbeitsnote] geben 83 [81] Prozent der Schüler die korrekte Notenstufe an, 5 [6] Prozent berichten eine niedrigere, 12 [13] Prozent eine höhere Note. In 1 [1] Prozent der Fälle beträgt die Unterschätzung mehr als eine Notenstufe, in 2 [3] Prozent der Fälle beträgt die Überschätzung mehr als eine Notenstufe.

Mittels Regressionen wurde überprüft, ob das Niveau der Akkuratheit (Differenz zwischen Lehrerangabe und selbst berichteter Note) durch die Faktoren Geschlecht, Leistungshöhe oder Fähigkeitsselbstkonzept vorhergesagt werden konnte. Kriterium in der ersten Regression war der Differenzwert für die Zeugnisnote, Prädiktoren waren Geschlecht (weiblich [-0.5],

männlich [0.5]), Leistungswert im Test (standardisiert), Fähigkeitsselbstkonzept (standardisiert) sowie die entsprechenden Produktterme. Das R der Regression war statistisch signifikant von Null verschieden, $F(7, 838) = 2.30$. Die Varianzaufklärung war mit einem R^2 von .02 [korrigiert .01] gering. Höhere Fähigkeitsselbstkonzepte waren in geringem Ausmaß assoziiert mit stärkerer Überschätzung der eigenen Note (standardisiertes $\beta = .10$, $p < .01$). Auch der Interaktionsterm Geschlecht X Fähigkeitsselbstkonzept hatte mit $\beta = .08$, $p < .05$ ein statistisch signifikantes, jedoch geringes Regressionsgewicht. Die grafische Auswertung der Interaktion (vgl. Aiken & West, 1991) zeigte, dass der Zusammenhang zwischen Fähigkeitsselbstkonzept und Akkuratheit bei Jungen enger war als bei Mädchen. Verwendet man in der Regression anstatt der Leistungsgruppierung die Schulform (Hauptschule [-1], Realschule [0], Gymnasium [1]) sowie die entsprechenden Interaktionsterme als Prädiktoren, so zeigen sich analoge Befunde.

Eine Regression (Prädiktoren Geschlecht, Leistungshöhe, Fähigkeitsselbstkonzept sowie Interaktionsterme) wurde für den Differenzwert bei der Klausurnote als Kriterium gerechnet. Das R der Regression war statistisch nicht signifikant von Null verschieden, $F(7, 829) = 0.92$. Keiner der Prädiktoren hatte ein signifikantes Regressionsgewicht. Dies Bild bleibt unverändert, wenn man anstatt der Testleistung die Schulform als Prädiktor in der Regression aufnimmt.¹

Die Validität der selbst berichteten und vom Lehrer angegebenen Noten wurde geprüft, indem die Zeugnisnoten und die Testleistungen klassenweise z-standardisiert und anschließend korreliert wurden. Die (anhand der Reliabilität des Mathematiktests einfach minderungskorrigierte) Korrelation zwischen

selbstberichteter Zeugnisnote und Testleistung betrug $\underline{r} = -.40$, die zwischen Lehrerangabe der Note und Testleistung $\underline{r} = -.41$. Eine analoge Analyse für die Noten der letzten Klassenarbeit ergibt eine Korrelation von $-.38$ beim Selbstbericht und $-.36$ bei der Lehrerangabe. Die Unterschiede zwischen den Korrelationspaaren waren statistisch nicht signifikant (\underline{z} -Werte < 0.5).

4 Zusammenfassende Diskussion

Die Ergebnisse der vorliegenden Studie zeigen zusammenfassend, dass von Schülern selbst berichtete Noten erstens relativ eng mit den vom Lehrer angegebenen Noten korrelieren, dass zweitens das Ausmaß der Notenüberschätzung der Schüler durch potentielle Moderatoren kaum in praktisch nennenswertem Ausmaß vorhergesagt werden kann. Weiterhin ergaben sich keine Hinweise für die Hypothese, dass selbst berichtete und vom Lehrer angegebene Noten unterschiedliche Kriteriumsvalidität hinsichtlich eines externen Leistungsmaßes aufweisen.

Die Enge des Zusammenhangs zwischen den selbst berichteten Noten und Lehrerangaben entspricht dabei im wesentlichen den Befunden von Frucot und Cook (1994). Interessant ist, dass die Korrelationen zwischen selbstberichteter Note und Lehrerangabe für das Zeugnis und für die Klassenarbeit in vergleichbarer Höhe liegen, obwohl die Zeugnisvergabe zum Untersuchungszeitpunkt mit viereinhalb Monaten deutlich länger zurücklag als die Rückgabe der letzten Klassenarbeit. Dies ist möglicherweise durch die höhere Wichtigkeit der Zeugnis gegenüber der Klassenarbeitsnote zu erklären. Bei der Interpretation der Höhe der Korrelationen ist zu bedenken, dass möglicherweise auch die Lehrerangaben kein absolut reliabler Indikator der tatsächlichen Noten ist. So können sich in wenigen

Fällen beim Übertragen der Noten aus dem Buch der Lehrkraft auf die Codeliste Schreibfehler ergeben. Schätzen wir daher die Reliabilität der Lehrerangabe für Zeugnisnote und Klassenarbeit mit .95, so müssen die entsprechenden Korrelationen zwischen selbst berichteter Note und Lehrerangabe einfach minderungskorrigiert werden. Es ergeben sich dann Werte von .92 bei der Zeugnisnote und .90 bei der Note aus der Klassenarbeit. In jedem Fall verdeutlichen die vorliegenden Daten, dass Schülerangaben auch länger zurückliegender Noten sehr eng mit den über die Lehrkraft erfassten Leistungsmaßen korrelieren.

Eine hohe Korrelation zwischen selbst berichteter Note und Lehrerangabe impliziert dabei aber nicht eine perfekte Akkuratheit des Niveaus der erinnerten Note: hohe Korrelationen können auch bestehen, obwohl sich alle Schüler/-innen systematisch über- oder unterschätzen. In Übereinstimmung mit Befunden aus dem angloamerikanischen Raum zeigt sich in unserer Studie im Mittel die Tendenz, das eigene Leistungsniveau sowohl in der Klassenarbeit als auch im letzten Zeugnis zu überschätzen. Denkbar ist, dass Mechanismen des Self-Enhancements zu Erinnerungsverzerrungen führen oder aber sich die Schüler in der Befragung als leistungsstärker präsentieren wollen, als sie tatsächlich sind (impression-management, vgl. Shepperd, 1993). Das Ausmaß der Überschätzung ist jedoch mit einem Umfang von einer zehntel Standardabweichung nicht massiv. Zudem scheint für das Ausmaß der Überschätzung nicht von starkem Einfluss zu sein, ob die Note der kürzlich zurück gegebenen Klassenarbeit oder die Note aus dem letzten Zeugnis angegeben wird.

Für die Forschungspraxis von wesentlicher Bedeutung ist, dass sich für die Hypothese einer in Abhängigkeit von Modera-

toren differentiellen Überschätzung der Noten überwiegend keine überzeugenden Belege zeigen (wie bereits die Betrachtung der äußerst geringen Varianzaufklärung der Regressionsfunktionen zeigt). Ebenso wie sich in der Arbeit von Goldman et al. (1990) für den GPA in Abhängigkeit der Leistungsstärke kein unterschiedliches Akkuratheitsniveau zeigte, werden von Schülern die Mathematiknoten der letzten Klassenarbeit und des letzten Zeugnisses in Abhängigkeit vom Leistungsniveau nicht unterschiedlich überschätzt. Es ist zwar denkbar, dass Schüler nach Erhalt einer schlechten Note die Validität dieser stärker in Frage stellen als nach Erhalt einer guten Note (vgl. hierzu Shepperd, 1993), für die Annahme eines je nach Leistungsniveau unterschiedlichen Abweichens des Selbstberichts von der Lehrerangabe ergaben sich jedoch in unserer Studie keine Hinweise.

Die Stärke der Überschätzung der Zeugnisnote und Klassenarbeitsnote ist nicht abhängig vom Schülergeschlecht, womit unsere Befunde für Einzelnoten im deutschen Schulsystem denen der Arbeit von Frucot und Cook (1994) zum GPA entsprechen (vgl. jedoch Goldman et al., 1990). Einen statistisch signifikanten, aber geringen Effekt übte die Höhe des mathematischen Fähigkeitsselbstkonzepts auf das Niveau der Erinnerung der Zeugnisnote aus. Schüler/-innen mit hohem Fähigkeitsselbstkonzept zeigten eine stärkere Überschätzung als Schüler mit niedrigem Fähigkeitsselbstkonzept, dies scheint insbesondere für Jungen der Fall zu sein. Die Betrachtung der Stärke der β -Gewichte zeigt jedoch, dass das Ausmaß des Zusammenhangs nicht als substantiell zu bezeichnen ist; bei der Vorhersage der Akkuratheit der selbst berichteten Klassenarbeitsnote (Differenzwert) war er statistisch nicht signifikant.

Die Korrelationen sowohl der selbst berichteten wie der vom Lehrer angegebenen Note mit der über die TIMSS-Aufgaben erfassten Leistung unterscheiden sich statistisch nicht voneinander. Für die Behauptung unterschiedlicher kriterialer Validität selbst berichteter und vom Lehrer notierter Noten ergibt sich damit keine Stützung. Es ist einsichtig, dass die Korrelation bei dem aggregierten Maß Zeugnisnote höher ausfällt als bei der Klassenarbeit, da die TIMSS-Aufgaben das gesamte Spektrum des Curriculums abdecken (vgl. Baumert et al., 1998). Unsere Befunde liefern zusammenfassend keine Hinweise, dass Schülerangaben von Noten ein weniger valider Leistungsindikator sind als die entsprechenden Lehrerangaben. Selbst berichtete Leistungsangaben scheinen insgesamt ein ebenso brauchbares Leistungsmaß zu sein wie die entsprechenden Lehrerangaben. Unbenommen hiervon ist die Diskussion darüber, ob Noten generell einen geeigneten Leistungsindikator darstellen (zsf. Tent, 2001). Dies muss je nach Besonderheit der jeweiligen Untersuchung entschieden werden.

Literaturverzeichnis

Aiken, L.S. & West, S.G. (1991). Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA: Sage.

Baumert, J., Lehmann, R., Lehrke, M., Clausen, M., Hosenfeld, I., Neubrand, J., Patjens, S., Jungclaus, H. & Günther, W. (Hrsg.) (1998). Testaufgaben Mathematik. TIMSS 7./8. Klasse (Population 2) [Materialien aus der Bildungsforschung, Bd. 60]. Berlin: Max-Planck-Institut für Bildungsforschung

Dickhäuser, O. (2003). Überprüfung des erweiterten Modells des internal/external frame of reference. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 35, 200-207.

Dobbins, G. H., Farh, J.-L. & Werbel, J. D. (1993). The influence of self-monitoring on inflation of grade-point averages for research and selection purposes. Journal of Applied Social Psychology, 23, 321-334.

Frucot, V. G. & Cook, G. L. (1994). Further research on the accuracy of students' self-reported grade point averages, SAT scores, and course grades. Perceptual and Motor Skills, 79, 743-746.

Goldman, B. A., Flake, W. L. & Matheson, M. B. (1990). Accuracy of college students' perception of their SAT scores, high school and college grade point averages relative to their ability. Perceptual and Motor Skills, 70, 514.

Helmke, A. (1992). Selbstvertrauen und schulische Leistungen. Göttingen: Hogrefe.

Kirk, B. A. & Sereda, L. (1969). Accuracy of self-reported college grade averages and characteristics of non and discrepant responders. Educational and Psychological Measurement, 29, 147-155.

Möller, J., Streblow, L., Pohlmann, B. & Köller, O. (2003). An Extension to the Classical Internal/External Frame of Reference Model: Assimilation and Contrast Effects. Manuscript submitted for publication.

Schmela, M. (1993). Abiturienten erinnern sich an ihre alten Zeugnisnoten -- Zur Qualität leistungsbezogener autobiographischer Erinnerungen. Zeitschrift für Pädagogische Psychologie, 7, 47-58.

Shepperd, J.A. (1993). Student derogation of the Scholastic Aptitude Test: Biases in perceptions and presentations of college board scores. Basic and Applied Social Psychology, 14, 455-473.

Tent, L. (2001). Zensuren. In D.H. Rost (Hrsg.), Handwörterbuch Pädagogische Psychologie (2. Aufl., S. 805-811). Weinheim: Psychologie Verlags Union.

Zimmerman, M.A., Caldwell, C. H. & Bernat, D. H. (2002). Discrepancy between self-reports and school-record grade point average: Correlates with psychosocial outcomes among african american adolescents. Journal of Applied Social Psychology, 32, 68-109.

Fußnoten

¹Ein anonymes Gutachter wies darauf hin, dass die Verwendung von Differenzwerten in den Analysen problematisch sei. Dieses Kriterium wurde in den vorliegenden Analysen wegen der leichten Nachvollziehbarkeit der Befunde verwendet. Der gegen Differenzwerte häufig vorgebrachte Einwand der doppelten Messfehlerbehaftetheit dürfte im vorliegenden Fall nicht besonders stark ins Gewicht fallen, da die Lehrerangabe der Note vermutlich hoch reliabel gemessen ist (die Lehrer übertrugen die Noten aus ihrem Notenbuch). Alternativ zur Analyse von Differenzwerten bietet sich an, in einem ersten Schritt zunächst die selbst berichteten Noten (als Kriterium) aufgrund der Lehrerangaben (als Prädiktor) vorherzusagen und die Residuen dann in einem nächsten Schritt als Kriterium aufgrund von Geschlecht, Leistung (Testleistung bzw. Schulform), Fähigkeitsselbstkonzept vorherzusagen. Die beiden Regressionen für das Residuum bei der Zeugnisnote ergaben das gleiche Befundmuster wie für die Differenzwerte (jeweils schwache, statistische signifikante Regressionsgewichte des Fähigkeitsselbstkonzepts bzw. des Interaktionsterms Geschlecht X Fähigkeitsselbstkonzept). Die Regressionen für das Residuum bei der Klassenarbeitsnote ergaben ein schwaches Regressionsgewicht für den Prädiktor Fähigkeitsselbstkonzept (höhere Fähigkeitsselbstkonzepte sind assoziiert mit leichten Notenüberschätzungen) sowie ein schwaches Regressionsgewicht für die Schulform (in höheren Schulformen resultierten weniger Überschätzungen als in niedrigeren Schulformen). Alle anderen Prädiktoren einschließlich der Interaktionsterme wiesen wie bei der Analyse der Differenzwerte keine statistisch signifikanten Regressionsgewichte auf.