

**EFFEKTE KOMPETENZIELLEN FEEDBACKS
AUF PERFORMANZ (WISSENSCHAFTLICHES DENKEN),
MOTIVATION UND METAKOGNITION
VON LERNENDEN DER SEKUNDARSTUFE I**

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER PHILOSOPHISCHEN FAKULTÄT
DER CHRISTIAN-ALBRECHTS-UNIVERSITÄT
ZU KIEL

VORGELEGT VON

MAREIKE WOLLENSCHLÄGER

KIEL
(2012)

ERSTGUTACHTER: PROF. DR. JENS MÖLLER

ZWEITGUTACHTERIN: PROF. DR. UTE HARMS

DRITTGUTACHTER: PROF. DR. GABRIEL NAGY

TAG DER MÜNDLICHEN PRÜFUNG: 19.06.2012

DURCH DEN ZWEITEN PRODEKAN, PROF. DR. MICHAEL DÜRING ZUM DRUCK GENEHMIGT: 26.06.2012

DIE STUDIEN WURDEN VOM BMBF DURCH DAS RAHMENPROGRAMM ZUR FÖRDERUNG DER EMPIRISCHEN
BILDUNGSFORSCHUNG UNTERSTÜTZT UND IN ANBINDUNG AN DAS FORSCHUNGSPROGRAMM KOMDIF
(KOMPETENZMODELLE ALS BASIS FÜR EINE DIAGNOSEGESTÜTZTE INDIVIDUELLE FÖRDERUNG)
AN SCHULEN DES HAMBURGER SCHULVERSUCHS (ALLES>>KÖNNER) DURCHGEFÜHRT.

DANKSAGUNG

Zunächst möchte ich mich sehr herzlich bei meinen Betreuern und „Doktoreltern“ Prof. Ute Harms und Prof. Jens Möller für ihr effektives Feedback, ihre Begleitung auf dem Weg der Erkenntnisgewinnung bezüglich wissenschaftlicher Denk- und Arbeitsweisen sowie für ihre wohlwollende und kompetente Betreuung bedanken.

Ein Dank gilt auch meinen Kooperationspartnern Prof. Andrea Möller und Prof. Jürgen Mayer für die Bereitstellung der Experimentieraufgaben sowie Prof. John Hattie für seine ansteckende Begeisterung für das Forschungsthema „Feedback“ und die zeitintensive und produktive Zusammenarbeit während meines Forschungsaufenthalts.

An dieser Stelle sei auch den Schülerinnen und Schülern der an den Studien teilnehmenden Hamburger Schulen gedankt, ohne die die Dissertation nicht möglich gewesen wäre. Dasselbe gilt für ihre Lehrkräfte, hierbei besonders Astrid Nellner, die mit ihrem Organisationstalent und hohem zeitlichen Einsatz die Durchführung der Studien enorm erleichtert hat. Auch meinen wissenschaftlichen Hilfskräften Franziska Lahmann, Frederike Pohl und Ronja Just möchte ich herzlichst danken für die angenehme und vertrauensvolle Zusammenarbeit, die Unterstützung bei der Durchführung der Studien, für anregende Diskussionen in der Studienplanung und Ergebnisinterpretation sowie für das Lektorat englischer Manuskripte.

Ein großer Dank gilt auch Kerstin Schütte und Markus Lücken, die mich bei immer wieder auftauchenden statistischen Fragen unterstützten, sowie all meinen großartigen Kolleginnen und Kollegen, deren vielfältige und motivierende Unterstützung essentiell war.

Hierbei seien auch meine beiden Zimmerkolleginnen Sandra Nitz und Kirsten Redlin dankend erwähnt, die mir das Einleben in Kiel deutlich erleichterten und so manche Herausforderungen während der Promotionsphase hautnah miterlebten.

Auch Kristina Brandstädter und Kerstin Münchhoff möchte ich für die erfrischenden und vertrauensvollen Treffen, die guten Ratschläge und die freundschaftliche Begleitung sehr herzlich danken.

Für die Aufmunterung vor allem in der Endphase der Dissertation danke ich Nicolai Basel und Burkhard Schroeter, die in der tagtäglichen Zusammenarbeit durch ihren Humor erheblich zum Stressabbau beitrugen. Frederike Pohl, Sandra Nitz, Nicolai Basel, Jörg Großschedl und Jens Wollenschläger sei für das Korrekturlesen der Endfassung gedankt.

Von ganzem Herzen danke ich meinen Eltern Olly und Manfred Buck, die mir ein Vorbild waren und sind, die mich mit ihrer bedingungslosen Liebe begleitet und unterstützt und Träume ermöglicht haben. Auch meinen Geschwistern mit Familien, meinen Freunden und Patenkindern danke ich für die allgegenwärtige Unterstützung, für hilfreiche Telefonate (zu jeder Tages- und Nachtzeit) und aufmunternde Besuche sowie die Sicherheit, jederzeit aufgefangen zu werden.

Dankend erwähnen möchte ich auch Jens Christof Schreiber, dessen feine, charakterstarke und wertschätzende Persönlichkeit mein Leben zutiefst und fortwirkend bereichert hat.

Mein ganz besonderer Dank gilt meinem Ehemann Jens Wollenschläger, der mich mit seiner ausgleichenden, humorvollen, motivierenden und klugen Art beraten und einzigartig unterstützt hat, der alle Höhen und Tiefen (nicht nur die einer Dissertation, aber eben auch diese) bedingungslos mit mir durchstand und mit dem ich das kostbare Privileg habe, mein Leben teilen zu dürfen.

INHALTSVERZEICHNIS

EINLEITUNG

FEEDBACKFORSCHUNG	8
KOMPETENZFORSCHUNG	20
FORSCHUNGSFRAGEN DER VIER TEILSTUDIEN	26

STUDIE 1

EFFEKTE KOMPETENZIELLER RÜCKMELDUNG BEIM WISSENSCHAFTLICHEN DENKEN	43
--	----

STUDIE 2

COMPETENTIAL FEEDBACK EFFECTS ON PERFORMANCE: ARE THEY MEDIATED BY CALIBRATION?	63
---	----

STUDIE 3

EXTERNAL VERSUS INTERNAL COMPETENTIAL FEEDBACK: DIFFERENTIAL EFFECTS ON PERFORMANCE AND CALIBRATION?	97
---	----

STUDIE 4

IST KOMPETENZIELLES FREMDFEEDBACK ÜBERLEGEN, WEIL ES ALS EFFEKTIVER WAHRGENOMMEN WIRD?	133
---	-----

DISKUSSION UND AUSBLICK

DISKUSSION DER VIER TEILSTUDIEN	160
ANSCHLUSSFRAGESTELLUNGEN ZUKÜNFTIGER FORSCHUNG.....	164
GESAMTFAZIT.....	168

EFFEKTE KOMPETENZIELLEN FEEDBACKS AUF PERFORMANZ (WISSENSCHAFTLICHES DENKEN), MOTIVATION UND METAKOGNITION VON LERNENDEN DER SEKUNDARSTUFE I

Die vorliegende Dissertation umfasst vier empirische Studien, die allesamt Effekte eines auf einem Kompetenzmodell basierenden Feedbacks, hier bezeichnet als kompetenziell, im schulischen Kontext in der Domäne wissenschaftliches Denken im naturwissenschaftlichen Unterricht untersuchen. Das einleitende Kapitel stellt eine theoretische Hinführung zur Definition des kompetenziellen Feedbacks dar, verdeutlicht dabei die der Planung der Studien zugrundeliegenden theoretischen und empirischen Befunde und mündet in die Darstellung der Ziele jeder Teilstudie. Da kompetenzielles Feedback an die Feedback- und Kompetenzforschung anschließt, wird auf diese zwei Forschungsbereiche näher eingegangen.

1 FEEDBACKFORSCHUNG

Das folgende Kapitel umfasst die Definition von Feedback, die Darstellung von Feedbacktheorien und -modellen und schließlich die Darstellung der für die Studien relevanten empirischen Befunde im Bereich der Feedbackforschung.

1.1 Definition von Feedback

Feedback wird generell definiert als Information, die einer Person während oder nach einem Prozess oder Prozessschritt rückgemeldet wird, um regulierend auf Prozess- oder Prozessschritt-Ebene einzuwirken (vgl. Narciss, 2006, 2008). Feedback im Lehr-/Lernkontext

(im Sinne von Leistungsrückmeldungen) ist als Information definiert, welche an der Lücke zwischen einem aktuellen und einem angestrebten Lernziel ansetzt und diese schließen soll (Hattie, 2009; Sadler, 1989). Somit entspricht Feedback einer *post-response information* im lernpsychologischen Sinn. In der Psychologie wurde schon früh die Wirkung von Feedback auf Lehr-/Lernsituationen untersucht (Kluger & DeNisi, 1996; Kulhavy & Wager, 1993). Thorndike (1913) beschrieb in seinem *law of effect*, dass das Ergebnis eines Verhaltens und die damit verknüpften Konsequenzen zukünftiges Verhalten beeinflussen. Somit bedingen die Informationen über Verhaltensergebnisse und -konsequenzen, die im Feedback enthalten sind, ob das gezeigte Verhalten beibehalten oder verändert wird. Feedback kann dabei entweder die Funktion als Motivator, Anreiz oder auch als Informationszugewinn bezüglich des gezeigten Verhaltens, zum Beispiel bezüglich einer gegebenen Antwort umfassen, sodass diese dann vom Lernenden verändert oder validiert werden kann (Kulhavy & Wager, 1993).

1.2 Feedbacktheorien und -modelle

Im Folgenden werden drei Feedbacktheorien vorgestellt, welche die theoretische Basis für die Ableitung der Hypothesen der vier durchgeführten Studien darstellen. Die Theorien zum Wirkmechanismus von Feedback sind in Modellen veranschaulicht und betrachten unterschiedliche Aspekte von Feedback. Zunächst wird ein Modell zur Ableitung der Feedbackeffektivität vorgestellt (Hattie & Timperley, 2007), dann ein Modell, welches sowohl Kognition als auch Motivation im Feedbackprozess beachtet (Vollmeyer & Rheinberg, 1998), und schließlich ein Modell, welches zwischen Fremd- und Selbstfeedback (also nach der Quelle von Feedback) unterscheidet und deren Auswirkungen theoretisch skizziert (Butler & Winne, 1995). Die Darstellung der Theorien endet jeweils mit einem Ausblick, wie die vier durchgeführten Studien an diese Theorien anknüpfen.

1.2.1 Modell zur Feedbackeffektivität

Hattie und Timperley (2007) beschreiben den Wirkmechanismus von Feedback mit drei Feedbackfragen auf vier Feedbackebenen. Effektives Feedback sollte demnach die Fragen „*Where am I going?*“ (Was ist das Lernziel?), „*How am I going?*“ (Wie hat der oder die Lernende in Relation zum Lernziel abgeschnitten?) und „*Where to next?*“ (Wie kann das Lernziel zukünftig besser erreicht werden?) beantworten, um die Diskrepanz zwischen aktuellem und erwünschtem Lernziel zu reduzieren (Hattie & Timperley, 2007, S. 87). Basierend auf der Forschung von Kulhavy und Stock (1989), die zwischen aufgabenbezogenem (z.B. Darstellung der korrekten Lösung), instruktionsbezogenem (erneute Erläuterung des Aufgabenziels) und extrainstruktionalem Feedback (zusätzliche Informationen bezüglich Aufgabe und/oder Aufgabenziel) unterscheidet, differenzieren Hattie und Timperley (2007) vier Feedbackebenen: Aufgabenebene, Prozessebene, Selbstregulationsebene und Selbstebene. Feedback auf der Aufgabenebene bezieht sich auf die entsprechende Aufgabe und stellt Informationen bereit, wie gut die Aufgabe bearbeitet wurde. Feedback auf der Prozessebene bezieht sich auf den Prozess der Aufgabenbearbeitung und stellt für die Ziele und Bearbeitungsstrategien der Aufgabe betreffende Informationen bereit. Auf der Selbstregulationsebene zielt Feedback auf Strategien des Lernenden hinsichtlich Unterstützung, Überwachung und Handlungskontrolle zur Zielerreichung ab. Feedback auf der Selbstebene bewertet die Lernenden selbst (z.B. durch Lob) und beinhaltet keine Informationen bezüglich der Aufgabe, der Aufgabenbearbeitung oder der Strategien bei der Aufgabenbearbeitung. Feedback auf der Selbstebene ist dadurch nicht in der Lage, die Feedbackfragen zu beantworten (Hattie & Gan, 2011). Somit beschreiben Hattie und Wollenschläger (2012) in ihrer *multidimensional conceptualization of feedback* die Wirksamkeit der drei Feedbackfragen auf den drei Feedbackebenen: Aufgaben-, Prozess- und Selbstregulationsebene (siehe Abbildung 1).

Bei der theoretischen Verortung und der Ableitung der Hypothesen bezüglich der in den vorliegenden Studien untersuchten Feedbackart (kompetenzielles Feedback) wurde an den drei Feedbackfragen (*Where am I going?*, *How am I going?* und *Where to next?*) nach Hattie und Timperley (2007) angeknüpft. Bei der Konstruktion des zu untersuchenden Feedbacks wurde beachtet, dass die drei Feedbackfragen auf Aufgabenebene beantwortet wurden.

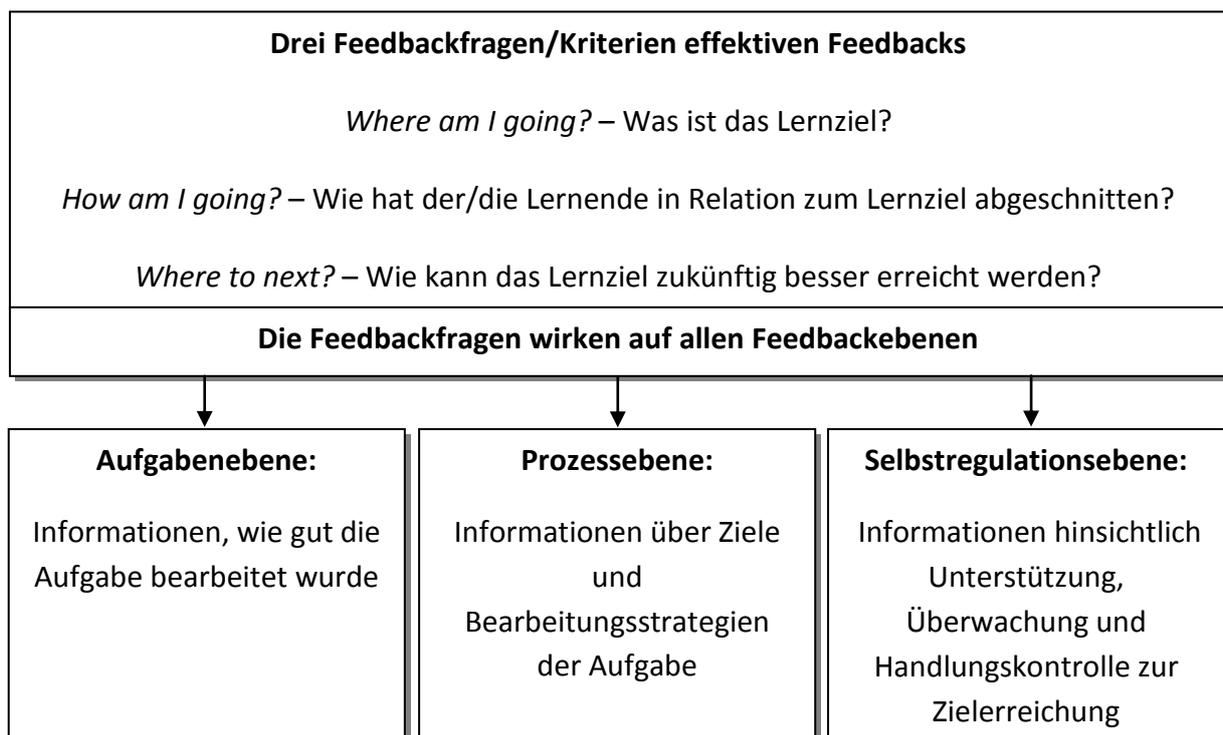


Abbildung 1: Modell der Feedbackeffektivität nach Hattie und Timperley (2007), angepasst nach Hattie und Wollenschläger (2012)

1.2.2 Kognitiv-motivationales Feedback-Modell

Obwohl das *cognitive-motivational process model* (Vollmeyer & Rheinberg, 1998) nicht explizit zur Beschreibung der Wirksamkeit von Feedback erstellt wurde, so bietet es doch einen theoretischen Rahmen zur Beschreibung und Ableitung kognitiver wie auch motivationaler Feedbackeffekte (vgl. Vollmeyer & Rheinberg, 2005). Dem Modell liegen einerseits Motivationstheorien (z.B. Anderson, 1993; Atkinson, 1957) und andererseits

theoretische Überlegungen zu Feedbackeffekten (*feedback intervention theory*, Kluger & DeNisi, 1996) zugrunde. Im *cognitive-motivational process model* wird angenommen, dass die Ausgangsmotivation (z.B. wahrgenommene Herausforderung, Interesse, wahrgenommene Erfolgswahrscheinlichkeit) den Lernprozess mediiert über die Strategiesystematik (Ausmaß des systematischen Umgangs von Lernenden mit der Aufgabe) und über die Motivation während der Aufgabenbearbeitung beeinflusst. Dieses Modell wurde auf die Feedbackforschung übertragen (vgl. Vollmeyer & Rheinberg, 2005), wobei Feedback dann die Rolle der Ausgangsmotivation einnimmt und direkt auf Strategiesystematik und Motivation während der Aufgabenbearbeitung einwirkt.

Folglich kann die Motivation nicht nur als unabhängige, sondern zugleich als abhängige Variable im Lernprozess und somit auch als abhängige Variable von Feedbackeffekten betrachtet werden. Hieran anschließend konzentrieren sich die vorliegenden Studien (Studien 1 und 2) auf die Motivation als abhängige Variable von Feedbackeffekten.

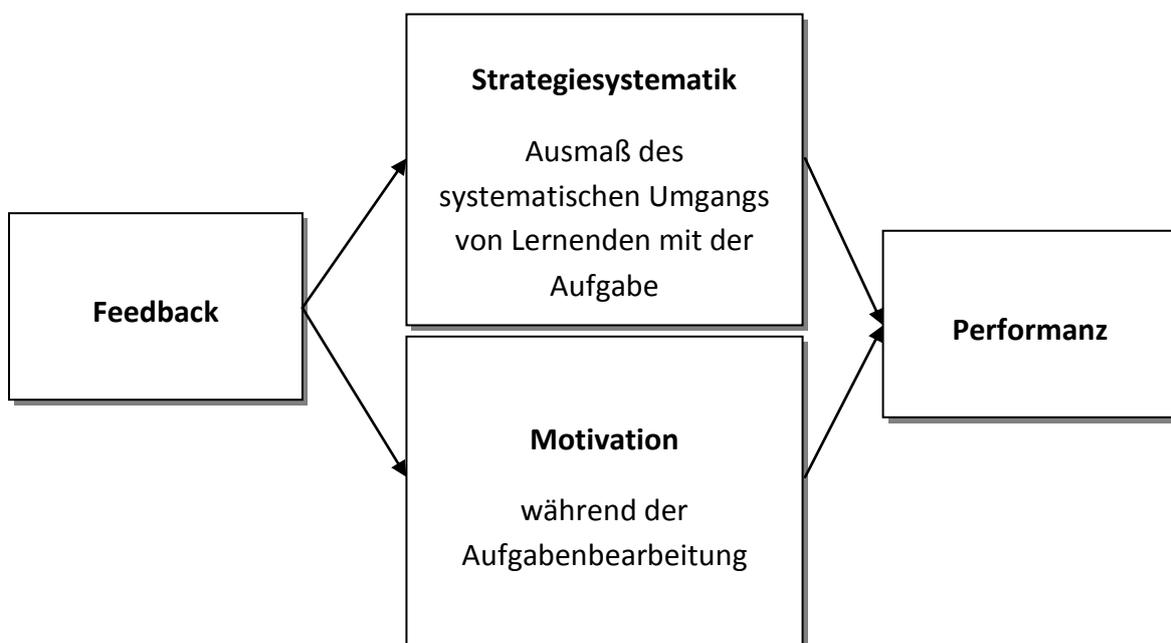


Abbildung 2: *Cognitive-motivational process model*, vereinfacht nach Vollmeyer und Rheinberg (1998)

1.2.3 Fremd- versus Selbstbewertung von Feedback

In Feedback-Definitionen wird Feedback oftmals von Verhaltenskonsequenzen abgegrenzt, die eine Person selbst wahrnimmt; Feedback wird demnach durch die Stimuli definiert, die kontingent während oder nach der Bearbeitung von Aufgaben dargeboten und die von einer externen Informationsquelle gezielt beeinflusst werden (vgl. Hattie & Timperley, 2007). Dieser unidimensionalen Sicht von Feedbacksender und -empfänger widersprechen Butler und Winne (1995) in ihrem *model of self-regulated learning*, in dem sie interne und externe Feedbackquellen innerhalb des selbstregulierten Lernens differenzieren. Externes Feedback entspricht dem von außen einwirkenden Feedback; internes Feedback findet laut Butler und Winne (1995) innerhalb des kognitiven Systems des Lernenden statt und beeinflusst gezielt dessen Wissen, Zielsetzungen und Strategien.

Im *model of self-regulated learning* (Butler & Winne, 1995) können zwei Dinge abgeleitet werden: zum einen die Differenzierung zwischen externem und internem Feedback, zum anderen, dass Feedbackeffekte im Kontext möglicher Interaktionen von Aufgabenanforderungen, instruktionalen Kontexten und Lernervoraussetzungen entstehen (Shute, 2008). Hiervon ausgehend untersuchen die durchgeführten Studien (Studien 3 und 4) gezielt Feedbackeffekte a) vergleichend zwischen Fremd- und Selbstbewertung und b) unter Einbezug möglicher zugrundeliegender mediierender Variablen.

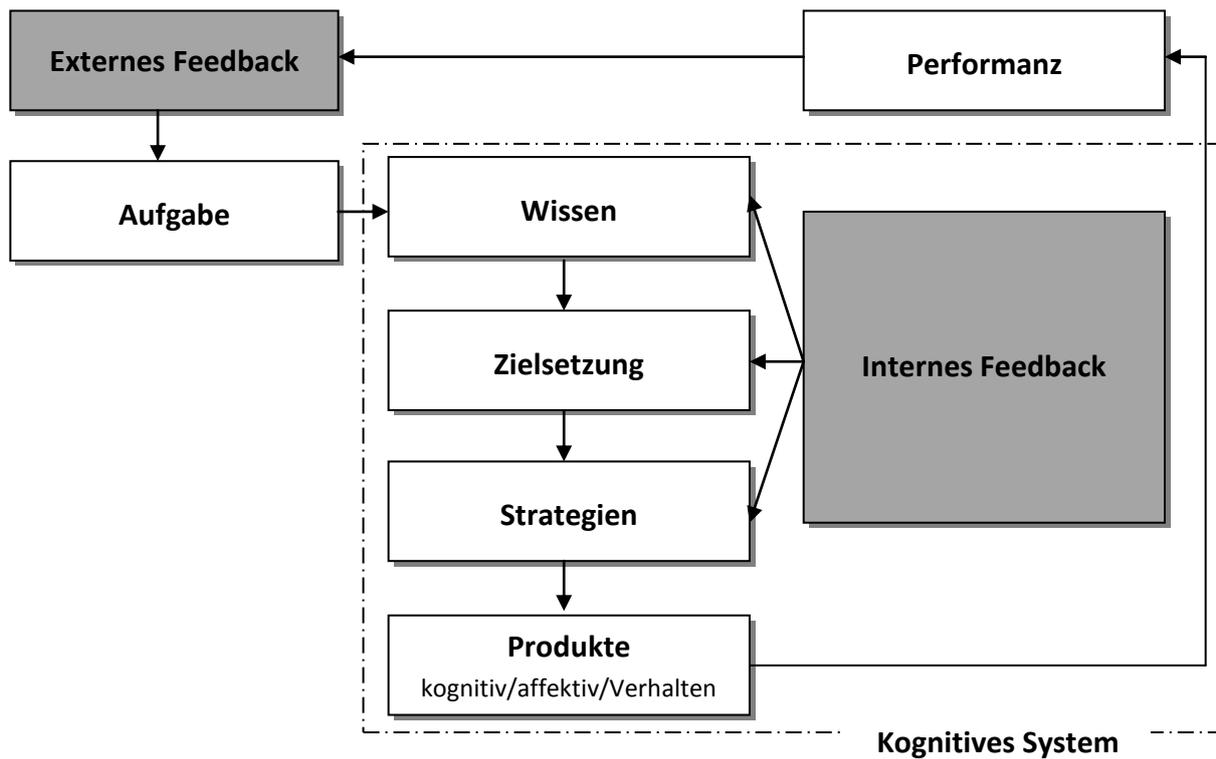


Abbildung 3: Model of self-regulated learning, vereinfacht nach Butler und Winne (1995)

Zusammenfassend wird deutlich, dass bei der theoretischen Fundierung von Feedback sowohl personale Faktoren wie kognitive, motivationale, metakognitive (z.B. Strategien) als auch situationale (z.B. Fremd- vs. Selbstfeedback) Faktoren betrachtet werden und somit auch bei der Wirkung von Feedback untersucht werden sollten (vgl. Hattie & Wollenschläger, 2012).

1.3 Empirische Befunde

Im Folgenden werden empirische Befunde aus der Feedbackforschung präsentiert, welche die empirische Basis für die Ableitung der Hypothesen der vier durchgeführten Studien darstellen. Die Ergebnisdarstellung der Unterkapitel endet jeweils mit einem Ausblick auf offene Forschungsfragen des jeweiligen Forschungsbereichs.

1.3.1 Empirische Befunde zur Feedbackeffektivität auf die Performanz von Lernenden

In Zusammenfassungen über den Stand der Feedbackforschung innerhalb der Lehr-/Lernforschung (Black & Wiliam, 1998; Mory, 2004; Shute, 2008) wie auch in Meta-Analysen (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie, 2009, 2012; Kluger & DeNisi, 1996; Kulhavy & Wager, 1993) wird Feedback im Unterricht (vgl. Hattie & Timperley, 2007) als essenzielle Einflussgröße von effektiven Lehr-/Lernprozessen identifiziert. Allerdings zeigt sich eine inkonsistente Befundlage, was die Effektivität von Feedback auf die Performanz betrifft; Effektstärken weisen eine hohe Variation auf, und nicht jede Form des Feedbacks scheint sich positiv auf den Lernprozess auszuwirken (Bangert-Drowns et al., 1991; Hattie, 2009; Kulhavy, 1977).

Zusammenfassend zeigt sich in den Meta-Analysen (Bangert-Drowns et al., 1991; Hattie, 2009; Shute, 2008) eine positive Auswirkung von Feedback auf die Performanz von Lernenden, wenn das Feedback

- (1) eine Lücke zwischen Ist- und Sollzustand verdeutlicht,
- (2) ungünstige Strategien bei der Aufgabenlösung verändert sowie
- (3) Informationen vor allem hinsichtlich der Frage enthält, wie die Aufgabebearbeitung zukünftig verbessert werden kann. Diese drei Kriterien effektiven Feedbacks finden sich in den drei Feedbackfragen von Hattie und Timperley (2007) wieder: *Where am I going?*, *How am I going?* und *Where to next?*

Es besteht Forschungsbedarf hinsichtlich der Untersuchung des auf diesen Feedbackfragen (Hattie & Timperley, 2007) basierenden Feedbacks und dessen Effekten vor allem im Schulkontext (Hattie, 2009).

1.3.2 Empirische Befunde zur Feedbackeffektivität auf die Motivation von Lernenden

Es existieren weitaus mehr Studien zu den Effekten von Feedback auf die Performanz als auf die Motivation von Lernenden (vgl. Hattie, 2009; Rakoczy, Klieme, Bürgermeister, & Harks, 2008; Vollmeyer & Rheinberg, 2005). Trotz der schwierigeren Erfassung motivationaler Variablen (Vollmeyer & Rheinberg, 2005) können positive Feedbackeffekte auf Motivation nachgewiesen werden (z.B. Rakoczy et al., 2008). Vor allem im Bereich der Forschung zu Bezugsnormen (vgl. Rheinberg, 1980, 2008) sind motivationale Effekte vielfach untersucht (Mischo & Rheinberg, 1995; Köller, 2005a,b). So kann gezeigt werden, dass die individuelle Bezugsnorm (Vergleich der Performanz eines Lernenden mit einem vorhergehenden Leistungsergebnis desselben Lernenden) vor allem für lernschwächere Schülerinnen und Schüler von Vorteil ist (Mischo & Rheinberg, 1995). Kluger und DeNisi (1996) weisen ein Defizit normativen Feedbacks, d.h. eines Vergleichs der Einzelleistung des Lernenden mit den Leistungen der Mitschülerinnen und Mitschüler, nach. Diese Form des Feedbacks, auch „soziale Bezugsnorm“ genannt (Rheinberg, 2008, S. 179), führt signifikant häufiger dazu, dass leistungsschwächere Schülerinnen und Schüler ihre niedrige Leistung auf ihre mangelnde Fähigkeit attribuieren, eine niedrigere Leistung bei zukünftiger Aufgabebearbeitung erwarten und geringer motiviert sind, was die Bearbeitung zukünftiger Aufgaben betrifft (Rheinberg, 1980, 2008). Eine kriteriale oder „sachliche“ Bezugsnorm (Vergleich der Performanz eines Lernenden mit einem im beabsichtigten Handlungszweck liegenden Standard wie beispielsweise dem Erwartungshorizont der Lehrkraft bezüglich der gestellten Aufgabe, Rheinberg, 2008, S. 179) wurde bisher noch nicht bezüglich Motivationseffekten vergleichend zu anderen Bezugsnormen untersucht (vgl. Rheinberg, Duscha, & Michels, 1980). In der Literatur wird die Hypothese formuliert, dass eine kriteriale Bezugsnorm durch das Aufzeigen der noch zu erreichenden Lernziele möglicherweise mit Motivationseinbußen einhergehe, da die Lernenden dieses Feedback als „Defizitmeldung“ verstehen könnten

(Rheinberg, 2008, S. 180). Diese Hypothese steht im Widerspruch zu den Kriterien effektiven Feedbacks (Hattie & Timperley, 2007), insbesondere zur zweiten Feedbackfrage: *How am I going?* (Wie hat der oder die Lernende in Relation zum Lernziel abgeschnitten?). Aufgrund dieses Kriteriums effektiven Feedbacks könnte konträr zu der in der oben genannten Hypothese angenommen werden, dass kriteriales Feedback motivationsförderlich ist.

Es besteht folglich weiterer Forschungsbedarf bezüglich einer systematischen Überprüfung von Feedbackeffekten auf die Motivation von Lernenden, wenn dem Feedback eine kriteriale Bezugsnorm zugrundeliegt.

1.3.3 Empirische Befunde zur Selbstbewertung von Feedback

Basierend auf Butler und Winne (1995) wird in der Feedbackforschung auch die Selbstbewertung der eigenen Leistung (internes Feedback) untersucht. Der Prozess der Selbstbewertung der eigenen Leistung wird in der Literatur zu formativem Assessment als zentraler Wirkfaktor der Performanzsteigerung durch Feedback benannt (z.B. Bennett, 2011; Black & Wiliam, 1998; Sadler, 1998). Es liegen zahlreiche Forschungsergebnisse und auch Meta-Analysen zur Untersuchung der Effekte von Selbstbewertung auf die Performanz von Lernenden auf Hochschulebene vor (Boud & Falchikov, 1989, 2006; Dochy, Segers, & Sluismans, 1999; Falchikov & Boud, 1989; Lew, Alwis, & Schmidt, 2010; Nulty, 2011); aber auch in der Grundschule (z.B. Towler & Broadfoot, 1992) und der Sekundarstufe 1 (z.B. White & Frederiksen, 1998) konnte gezeigt werden, dass Selbstfeedback die Performanz von Lernenden günstig beeinflusst. So wird beispielsweise angenommen, dass Lernende sich selbst lernförderliches Feedback geben können, wenn sie dafür angemessene Bedingungen und Unterstützung erhalten (Goodrich Andrade & Boulay, 2003). Basierend auf dieser Hypothese untersuchten Goodrich Andrade und Boulay (2003) Effekte von Selbstbewertungen der eigenen Leistung anhand vorgegebener Rubriken. Die vorgegebenen

Rubriken enthielten die zu erreichenden Lerninhalte. Durch die Selbstbewertung der eigenen Leistung in Relation zu den Rubriken konnten die Lernenden ihre erreichten und noch zu erreichenden Lerninhalte evaluieren. Diese sogenannten *standards-referenced assessment tools* verbesserten die Performanz von Lernenden im Bereich der Schreibleistung im Vergleich zu einer Kontrollgruppe ohne Selbstbewertungsintervention (vgl. Goodrich Andrade & Boulay, 2003).

Bisher wurden Effekte von Selbstbewertungen, auf einem vorher definierten Kriterium basierend, mit einer Kontrollgruppe ohne Selbstbewertungsintervention verglichen (vgl. Andrade & Du, 2007; Andrade, Du, & Wang, 2008). Es besteht Forschungsbedarf im Hinblick auf eine vergleichende Untersuchung von Feedbackeffekten, wenn Selbst- und Fremdfedback dasselbe Kriterium zugrundeliegt.

1.3.4 Empirische Befunde zur Mediation von Feedbackeffekten

Wie Butler und Winne (1995) im *model of self-regulated learning* verdeutlichen, sind vermittelnde Variablen des Feedbackeffekts auf die Performanz von großer Bedeutung (Hattie & Gan, 2011). In neueren Forschungsarbeiten zu Feedbackeffekten wird dieser Ansatz aufgegriffen. Zur Betrachtung der Feedbackeffekte im Lernprozess wurden beispielsweise Lernervoraussetzungen (Labuhn, Zimmerman, & Hasselhorn, 2010) oder metakognitive und emotionale Variablen wie beispielsweise die kognitive Verarbeitungstiefe und die emotionale Erlebensqualität (Rakoczy et al., 2008) oder die wahrgenommene Nützlichkeit von Feedback (Harks, Rakoczy, Hattie, Klieme, & Besser, 2011) einbezogen. Auch in der Peer-Feedback-Forschung werden beispielsweise die Genauigkeit des Feedbacks (z.B. Gielen, Peeters, Dochy, Onghena, & Struyven, 2010) und die wahrgenommene Qualität des Feedbackgebers, d.h. die Wahrnehmung des Feedbackgebers als Autorität (z.B. Strijbos, Narciss, & Dünnebier, 2010), als Mediatoren der Feedbackeffektivität untersucht.

Es besteht weiterer Forschungsbedarf hinsichtlich der systematischen und standardisierten Untersuchung verschiedener Feedbacktypen (Rakoczy et al., 2008) und deren zugrundeliegender vermittelnder Variablen des Effekts von Feedback auf Performanz (Shute, 2008).

Zusammenfassend wird deutlich, dass im Bereich der Feedbackforschung zahlreiche Forschungsarbeiten vorhanden sind; allerdings besteht weiterer Forschungsbedarf u.a.

- 1) bei der Untersuchung von Feedback im Schulkontext, welches alle Kriterien effektiven Feedbacks erfüllt,
- 2) bezüglich der Effekte eines kriterialen Feedbacks auf Performanz und Motivation von Lernenden,
- 3) bezüglich vergleichender Feedbackeffekte von Fremd- versus Selbstbewertung, welche beide auf denselben Kriterien effektiven Feedbacks basieren und
- 4) bezüglich der Interaktion dieses Feedbacks mit Charakteristika der Lernenden, d.h. der Mediationsprozesse zwischen Feedback und Performanz.

Diese Forschungsfragen werden in den vier Teilstudien der vorliegenden Dissertation aufgegriffen und mit Forschungsfragen aus dem Bereich der Kompetenzforschung verbunden.

2 KOMPETENZFORSCHUNG

Das folgende Kapitel umfasst die Definition von Kompetenz, die Skizzierung der Grundlagen und des Nutzens von Kompetenzmodellen sowie die ausführliche Darstellung des für die vorliegenden Studien relevanten Kompetenzmodells.

2.1 Definition von Kompetenzen

Im Kontext der zunehmenden Fokussierung auf Bildungsziele und deren Konkretisierung in den Bildungsstandards hat die Definition von Kompetenzen (Klieme & Leutner, 2006; Weinert, 2001) und die Entwicklung von Kompetenzmodellen in der deutschsprachigen pädagogisch-psychologischen und fachdidaktischen Forschung einen zentralen Platz eingenommen (z.B. Klieme, Hartig, & Rauch, 2008; Klieme, Bürgermeister, Harks, Blum, Leiß, & Rakoczy, 2010; Viering, Fischer, & Neumann, 2010) und zunehmend auch international Anschluss gefunden (vgl. Connel, Sheridan, & Gardner, 2003; Köller & Parchmann, 2012; Zilker, Holliday, Fischer, Kauertz, Lederman, & Lederman, 2010).

Kompetenzen wurden „als Systeme aus spezifischen, prinzipiell erlernbaren Fertigkeiten, Kenntnissen und metakognitiven Wissen definiert, die es erlauben, eine Reihe von Anforderungen in bestimmten Alltags-, Schul- und Arbeitsumgebungen zu bewältigen“ (Mayer, Grube, & Möller, 2008, S. 64). Die Erlernbarkeit und Domänenspezifität von Kompetenzen grenzt diese von allgemeinen kognitiven Fähigkeiten ab (z.B. Thorndike & Hagen, 1971); somit sind Kompetenzen durch pädagogische Interventionen oder Bildungsprozesse beeinflussbar (Klieme & Leutner, 2006).

Den meisten Forschungsarbeiten im Bereich der Kompetenzforschung liegt eine Fokussierung auf den kognitiven Aspekt von Kompetenzen zugrunde; Kompetenzen sind dann definiert als „kontextspezifische kognitive Leistungsdispositionen, die sich funktional

auf Situationen und Anforderungen in bestimmten Domänen im Sinne von spezifischen Lern- und Handlungsbereichen beziehen“ (Klieme & Leutner, 2006, S. 1).

Durch diese Fokussierung wird gewährleistet, dass Kompetenzen bei Lernenden durch die Güte von Aufgabenbearbeitungen beobachtbar sind (Fischer & Draxler, 2006) und sich effizient über Papier-Bleistift-Tests erfassen lassen, wie dies zum Beispiel für die naturwissenschaftlichen Fächer gezeigt wurde (Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010).

Die vorliegenden vier Studien knüpfen an die kognitive Kompetenzdefinition an und untersuchen Effekte von Feedback auf naturwissenschaftliche Kompetenzen von Lernenden, dabei speziell auf die Güte von Aufgabenbearbeitungen im Bereich der naturwissenschaftlichen Erkenntnisgewinnung.

2.2 Aufbau und Nutzen von Kompetenzmodellen

Kompetenzmodelle zielen darauf ab, die Struktur und die zentralen domänenspezifischen Fertigkeiten darzustellen und beschreiben die Stufen ihres Erwerbs (Klieme et al., 2003). Dabei lassen sich Kompetenzentwicklungsmodelle von Kompetenzstrukturmodellen abgrenzen (Schecker & Parchmann, 2006). Entwicklungsmodelle beabsichtigen, den „zeitlichen Verlauf und die innere Dynamik des Kompetenzerwerbs abzubilden“ (Bernholt, Parchmann, & Commons, 2009, S. 224). Allerdings sind in der bisherigen Kompetenzforschung kaum Kompetenzentwicklungsmodelle vorhanden und auch deren empirische Überprüfung steht noch aus (vgl. Bernholt, Neumann, & Nentwig, 2012; Eggert & Bögeholz, 2006; Hammann, 2004). In Abgrenzung dazu wird in Strukturmodellen keine Vorhersage über den zeitlichen Verlauf getroffen. Es wird davon ausgegangen, dass die für die betreffende Kompetenz ausgewählten Teilkompetenzen und deren Ausprägungen (Kompetenzstufen) in einem Kompetenzstruktur-

modell systematisch angeordnet und beschrieben werden können (Schecker & Parchmann, 2006). Dabei liegt die Ausgangshypothese zugrunde, dass sich Kompetenzen hierarchisch abbilden lassen und die Lösung einer schwierigeren Aufgabe eine höhere Kompetenz anzeigt (vgl. Schecker & Parchmann, 2006). Kompetenzstrukturmodelle veranschaulichen, ähnlich einem Koordinatensystem, auf den jeweiligen Achsen die domänenspezifischen Teilkompetenzen, definierte Anforderungsbereiche und meistens eine inhaltliche Anbindung (Bernholt et al., 2009). Die Teilkompetenzen entsprechen den Fertigkeiten, welche die Gesamtkompetenz ausmachen, die Anforderungsbereiche umfassen die schwierigkeitsgenerierende Komponente (z.B. Komplexitätsniveaus), und die inhaltliche Anbindung ist beispielsweise durch curriculare Vorgaben oder Basiskonzepte gegeben (Schecker & Parchmann, 2006). Bisherige Kompetenzmodelle weisen deutliche Unterschiede in den Kompetenzen und Teilkompetenzen auf (Dimensionierung) und variieren in der Graduierung der schwierigkeitsgenerierenden Komponente (Bernholt et al., 2009). Trotz der Variation im Vorgehen der Graduierung ist generell eine hierarchische Anordnung der Kompetenzstufen vorzufinden; diese bieten „eine Alternative zur willkürlichen Setzung von Leistungsmerkmalen auf einem Kontinuum“ (Klieme et al., 2003, S. 81).

Auch wenn Kompetenzen und Kompetenzmodelle durchaus kritisch diskutiert werden (Schecker, 2012), so definieren sie *learning outcomes* (Köller & Parchmann, 2012) und ermöglichen die Überprüfbarkeit des Erreichens dieser *learning outcomes*, was bisher vor allem auf Bildungssystemebene geschah (vgl. PISA). Allerdings bieten Kompetenzmodelle auch auf Individualebene, d.h. auf Ebene der Schülerinnen und Schüler, Anknüpfungspunkte; Kompetenzmodelle haben das Potenzial, auf individueller Ebene den Kompetenzstand jedes einzelnen Lernenden im Rahmen einer formativen Evaluation zu erfassen (Bernholt et al., 2009; Koeppen, Hartig, Klieme, & Leutner, 2008).

Die Tatsache der hierarchisch definierten, empirisch überprüfbaren Graduierung von domänenspezifischen Kompetenzen in Kompetenzstufen (Klieme et al., 2003) sowie die Möglichkeit, Kompetenzmodelle auch auf Individualebene zu nutzen, könnte auf den Bereich der Leistungsrückmeldungen im Unterricht übertragen werden: Lernziele und das Erreichen dieser Lernziele wären dann nicht mehr subjektiv gesetzt (d.h. abhängig von der Lehrperson), sondern wären objektiv festgelegt (Jäger, 2008).

Im Bereich der Kompetenzforschung wurde infolgedessen die Forschungsaufgabe genannt, konkrete Effekte eines Feedbacks über den individuellen Leistungsstand einer spezifischen Kompetenz zu untersuchen (Koeppen et al., 2008). Diese Forschungsfrage wurde in der vorliegenden Dissertation aufgegriffen und in der Domäne des wissenschaftlichen Denkens, basierend auf einem Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung, untersucht.

2.3 Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung

Wissenschaftliches Denken ist ein „domänenspezifischer, wissensbasierter und komplexer Problemlöseprozess, der durch spezifische Prozeduren charakterisiert ist und in dem auf methodisches Wissen zurückgegriffen wird“ (Mayer et al., 2008, S. 65). Im Rahmenkonzept wissenschaftsmethodischer Kompetenzen (Mayer, 2007, S. 178) ist wissenschaftliches Denken neben den „Charakteristika der Naturwissenschaften (*nature of science*)“ und „praktischen Arbeitstechniken (*practical work*)“ als eigenständiger Bereich „wissenschaftliche Erkenntnismethoden (*scientific inquiry*)“ aufgeführt. In verschiedenen Ansätzen sind die zentralen Prozeduren wissenschaftlicher Erkenntnismethoden beschrieben (z.B. Giere, Bickle, & Mauldin, 2006; Waddington, Nentwig, & Schanze, 2007). Darauf inhaltlich aufbauend leitet Mayer (2007, S. 181), basierend auf der Struktur des Problemlöseprozesses (interne Repräsentation eines Problems, Generieren eines Lösungsplans und Evaluation der

Ergebnisse), vier Dimensionen wissenschaftlicher Erkenntnismethoden ab: „Naturwissenschaftliche Fragen formulieren, Hypothesen generieren, Experimente planen sowie Daten analysieren/Schlussfolgerungen ziehen“ (siehe Abbildung 4).

Der Dimensionierung des Modells naturwissenschaftlicher Erkenntnisgewinnung liegen diese Komponenten (Fragestellung, Hypothese, Planung und Deutung) zugrunde (Mayer et al., 2008); sie konnten als vier eigenständige Teilkompetenzen sowie als einer übergeordneten Kompetenz zugehörig identifiziert werden (vgl. Grube, Möller, & Mayer, 2007; Mayer et al., 2008).

Was die Graduierung der Niveaustufen betrifft, so sind diesbezüglich verschiedene Vorgehensweisen in Kompetenzmodellen für die Naturwissenschaften gewählt worden: zum einen werden naturwissenschaftliche Kompetenzen beispielsweise nach zunehmender Komplexität von Alltagskonzepten zu wissenschaftlichen Konzepten (Bybee, 2002), nach steigender Anzahl an Elementen und den funktionalen Verknüpfungen zwischen diesen Elementen (vgl. Kauertz, 2008; Kauertz et al., 2010), nach zunehmender Systematisierung des Inhaltsbereiches (z.B. Hammann, 2004), nach hierarchischer Komplexität (Bernholt, 2010) oder durch post-hoc-Analysen aufgrund von Itemkennwerten (vgl. TIMSS und PISA) graduiert. Im Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008) wird eine theoretische a-priori-Graduierung „nach zunehmender Komplexität“ vorgenommen (Mayer et al., 2008, S. 66f.), und zwar über die Anzahl an Elementen und die funktionalen Verknüpfungen zwischen diesen Elementen: „1) Untersuchung eines Faktors, 2) Untersuchung von Zusammenhängen, 3) Kontrollierte Untersuchung auf Basis von Konzeptverständnis, 4) elaborierte Untersuchung allgemeiner Zusammenhänge und 5) Selbständiges Lösen von offenen Problemen“. Das Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung umfasst demnach vier Teilkompetenzen, untergliedert im

Sinne des Problemlöseprozesses und nach zunehmender Komplexität graduiert (siehe Abbildung 4).

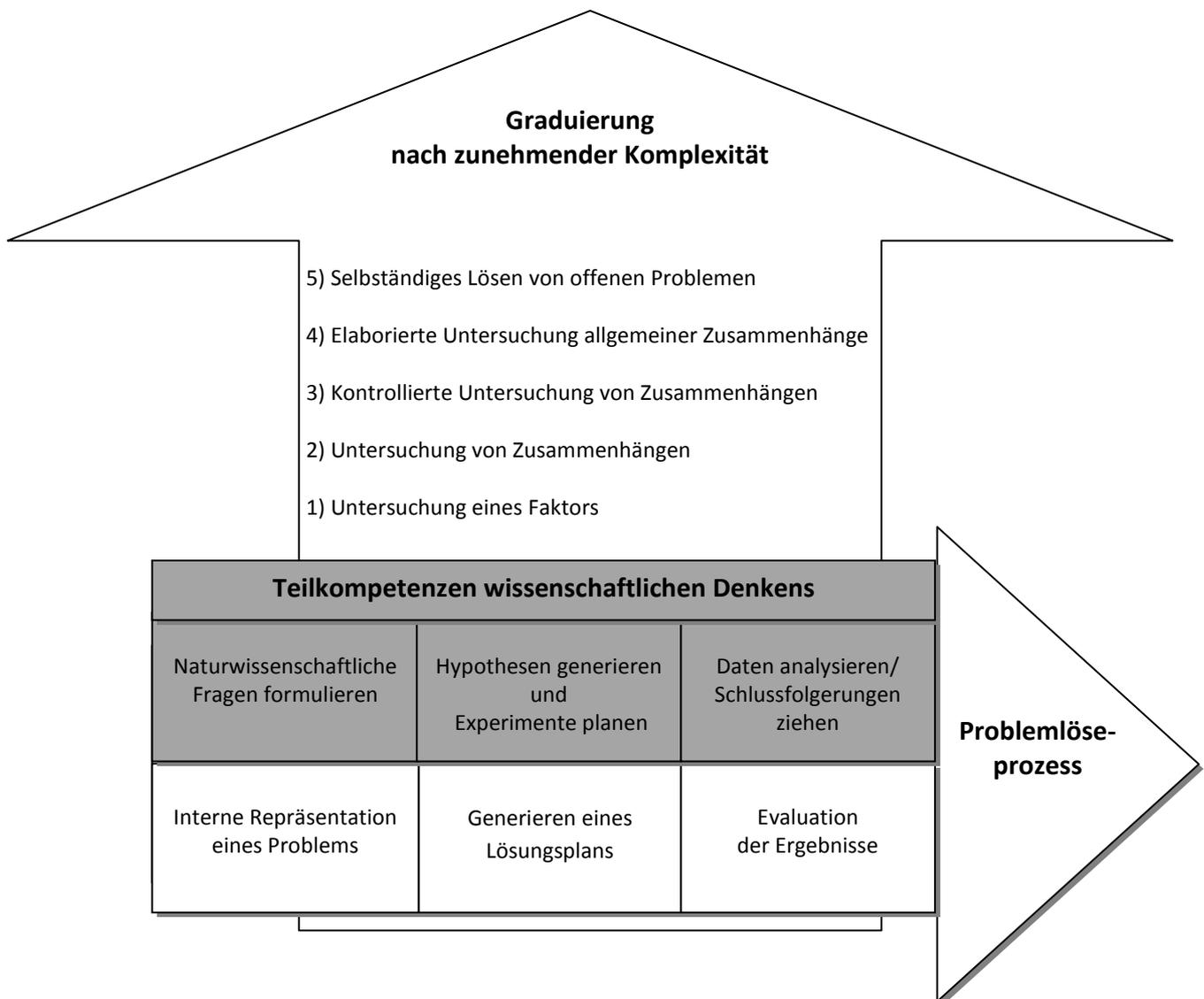


Abbildung 4: Schematische Darstellung der Dimensionierung und Graduierung (Kompetenzstufen 1-5) im Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008)

Die Graduierung nach zunehmender Komplexität konnte in der Validierungsstudie mit $N = 1553$ Schülerinnen und Schülern der Sekundarstufe 1 aufgezeigt werden (Mayer et al., 2008), allerdings war die Verteilung der Lernenden auf die Kompetenzstufen nicht normalverteilt. Vor allem die Kompetenzstufen 4 und 5 waren kaum besetzt; dies deutet auf

Aufgaben- und Auswertungsprobleme hin und erfordert eine weitere Überprüfung des Modells mit überarbeiteten Aufgaben (Mayer et al., 2008). Dies findet momentan implizit statt, da die Überlegungen von Mayer et al. (2008) in überarbeiteter Form in das ESNaS-Modell („Evaluation der Standards in den naturwissenschaftlichen Fächern der Sekundarstufe 1“, vgl. Walpuski, Kampa, Kauertz, & Wellnitz, 2008) einfließen.

Der vorliegenden Dissertation liegt das Ausgangsmodell zum wissenschaftlichen Denken von Mayer et al. (2008) zugrunde, da die durchgeführten Studien vor der Veröffentlichung des ESNaS-Modells im Jahre 2008 begonnen wurden, aufeinander aufbauen und vergleichbar sein müssen. Das Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung wird in den vier Teilstudien der Dissertation genutzt, um auf der theoretisch abgeleiteten und empirisch überprüften Dimensionierung und Gradierung ein theoretisch abgeleitetes und objektives Feedback entwickeln zu können.

3 ZIELE DER VIER TEILSTUDIEN

Nach der Darstellung des Gesamtziels der Dissertation, der Definition und Erläuterung des eingesetzten kompetenziellen Feedbacks folgt ein Überblick über die vier in der Dissertation enthaltenen Teilstudien sowie deren Forschungsfragen.

3.1 Gesamtziel der Dissertation und Definition kompetenziellen Feedbacks

Aufgrund der Tatsache, dass Kompetenzmodelle die theoretisch abgeleitete und empirisch überprüfte Dimensionierung und Graduierung einer domänenspezifischen Kompetenz beinhalten, diese Kompetenzmodelle auch auf Individualebene nutzbar gemacht werden sollen und Feedback in Relation zu in Kompetenzmodellen definierten Stufen bisher noch nicht untersucht wurde (siehe 2.2), untersuchen die vier Teilstudien der vorliegenden Dissertation Effekte eines auf einem Kompetenzmodell basierenden Feedbacks. Um zu

verdeutlichen, dass die Leistungsbeurteilung und -rückmeldung in Relation zu Stufen eines domänenspezifischen Kompetenzmodells stattfindet, wird das Feedback als kompetenziell bezeichnet.

Die Untersuchung eines auf einem Kompetenzmodell basierenden Feedbacks schließt nicht nur eine Forschungslücke im Bereich der Kompetenzforschung, sondern bearbeitet gleichzeitig die benannten Forschungsfragen aus der aktuellen Feedbackforschung (siehe 1.3.4):

Kompetenzielles Feedback

- 1) beinhaltet alle Kriterien effektiven Feedbacks (*Where am I going?, How am I going?* und *Where to next?*, Hattie & Timperley, 2007) und ermöglicht, diese im Schulkontext zu untersuchen.
- 2) ist kriterial, d.h. an einem externen, theoretisch fundierten und objektiven Standard (Kompetenzmodell) orientiert. Feedbackeffekte dieser kompetenziellen Erweiterung der kriterialen Bezugsnorm können nicht nur auf Performanz, sondern wie als Forschungsaufgabe identifiziert, auch auf Motivation von Lernenden überprüft werden.
- 3) ermöglicht die Untersuchung vergleichender Feedbackeffekte von Fremd- versus Selbstbewertung, welche beide auf denselben Kriterien (d.h. Stufen eines Kompetenzmodells) basieren.
- 4) bietet die Voraussetzung der systematischen Analyse von Mediationsprozessen von Feedbackeffekten auf Performanz.

3.2 Kompetenzielles Feedback zum wissenschaftlichen Denken

Die Umsetzung dieses theoretisch fundierten, innovativen Feedbacktyps (kompetenziell) findet in den vier Teilstudien, basierend auf dem in Abschnitt 2.3

dargestellten Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008), fokussiert auf die Teilkompetenz „Planung von Experimenten“, statt (siehe Abbildung 5).

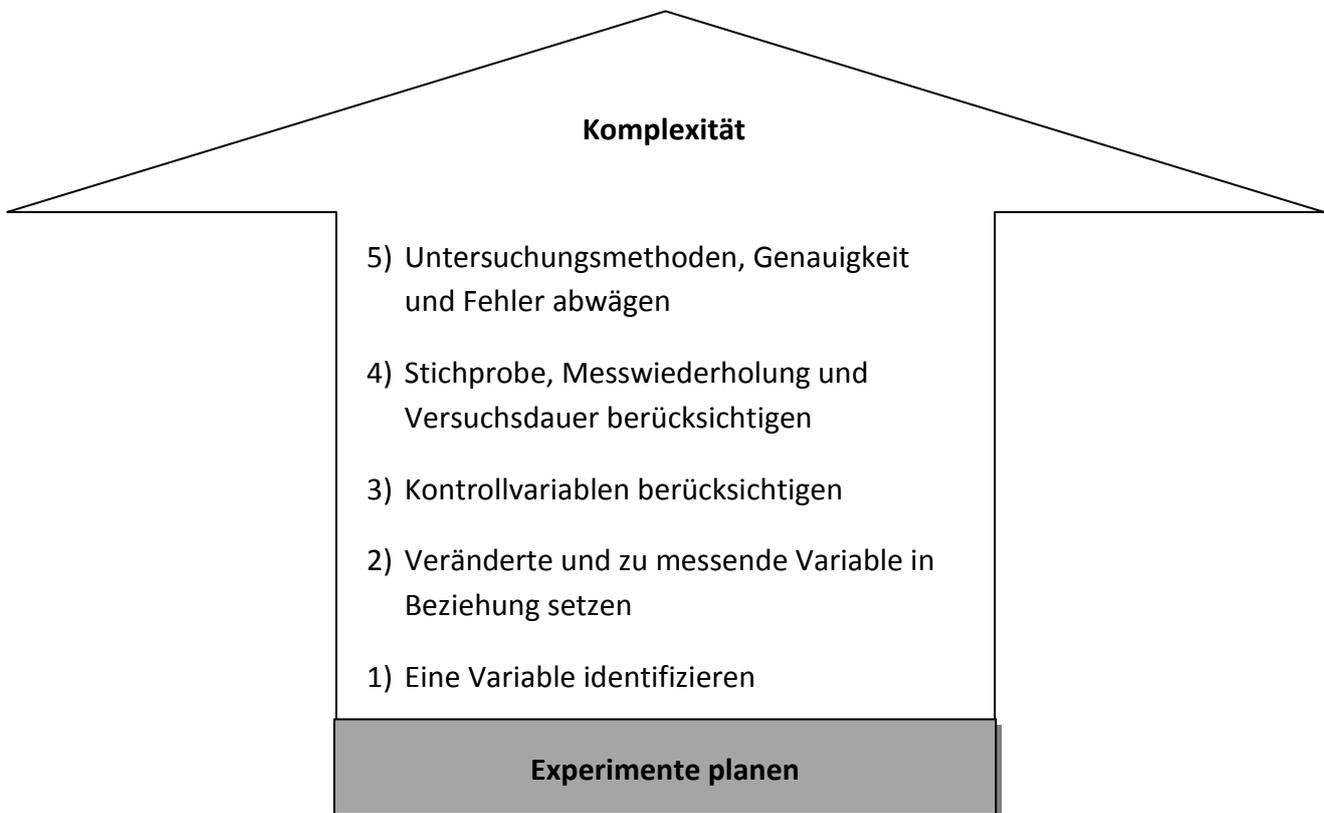


Abbildung 5: Graduierung (Kompetenzstufen 1-5) der Teilkompetenz „Experimente planen“ im Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008)

Das methodische Vorgehen der vier durchgeführten Studien der Dissertation umfasst zunächst die Übersetzung der Graduierung von Mayer et al. (2008) in eine schülergerechte Sprache. Anschließend kann den Lernenden auf dem in Abbildung 6 dargestellten Feedbackbogen rückgemeldet werden, was das Lernziel ist (Beachten aller aufgeführten Kompetenzstufen), wie der oder die Lernende in Relation zum Lernziel abgeschnitten hat (Ankreuzen der schon beachteten Kompetenzstufen) und wie das Lernziel zukünftig besser

erreicht werden kann (Aufmerksamkeitslenkung auf nichtangekreuzte Kompetenzstufen und inhaltliche Erklärung dieser Kompetenzstufen).

Die folgende Tabelle zeigt dir, welche Schritte du beachtet hast. Bei der Planung eines Experiments hast du ...	
... gesagt, was du untersuchen willst. Du hast eine Variable genannt. Eine Variable ist die zu untersuchende Bedingung, die im Experiment verändert, also variiert wird.	
... diese Variable verändert und gesagt, wie du die Veränderungen messen willst.	
... andere Einflussfaktoren, die du nicht untersuchen willst, gleichgehalten, d.h. nicht variiert.	
... gesagt, an wie vielen Untersuchungsgegenständen, wie oft oder wie lange du das Experiment durchführen willst.	
... über mögliche Fehler oder Schwächen in deiner Planung nachgedacht und hast dann eine neue Untersuchung entworfen.	

Abbildung 6: Kompetenzieller Feedbackbogen zum wissenschaftlichen Denken (Teilkompetenz Experimente planen), basierend auf Mayer et al. (2008); absteigend nach zunehmender Komplexität (Kompetenzstufen 1-5) angeordnet

3.3 Forschungsfragen der vier Teilstudien

Die vier Teilstudien der vorliegenden Dissertation vergleichen kompetenzielles Feedback mit einer Kontrollgruppe ohne Feedback (Studien 1 und 2) sowie mit anderen Feedbackarten wie sozialem Feedback (Studie 2) und internem Feedback (Studien 3 und 4). Außerdem werden Effekte kompetenziellen Feedbacks auf mehrere abhängige Variablen untersucht: auf die Performanz von Lernenden in der Domäne des wissenschaftlichen Denkens bezogen auf den Teilbereich der Planung von Experimenten (Studien 1 bis 4), auf deren Motivation (Studien 1 und 2), auf die Genauigkeit der Selbsteinschätzung der Lernenden (*calibration accuracy*, vgl. Schraw, Potenza, & Nebelsick-Gullet, 1993, Studien 2

und 3) und auf die wahrgenommene Feedbackeffektivität (Studie 4). Außerdem werden Mediationsprozesse des Feedbackeffekts auf Performanz untersucht (Studien 2-4).

Einen schematischen Überblick über die Studien gibt Abbildung 7:

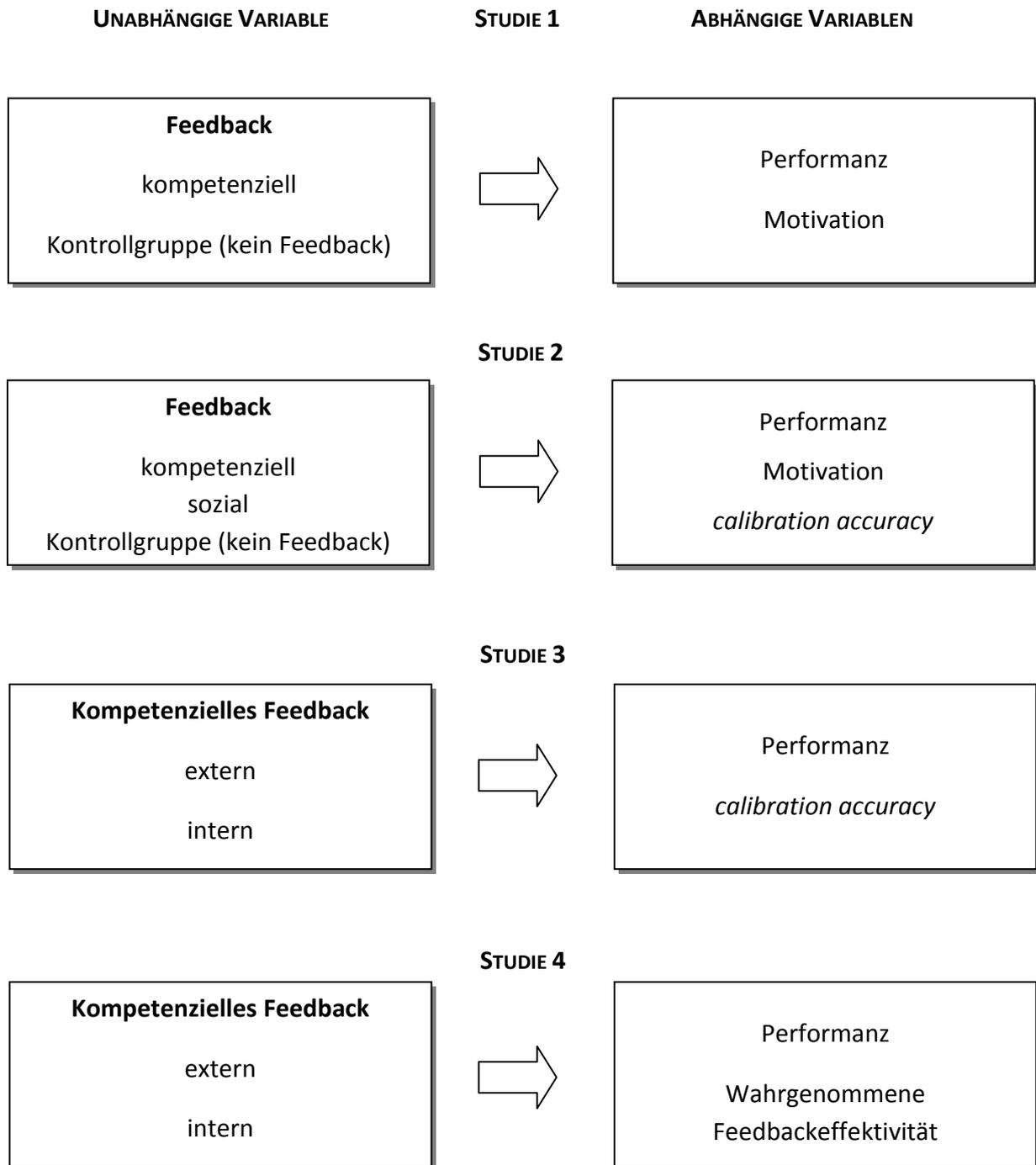


Abbildung 7: Schematische Darstellung der unabhängigen und abhängigen Variablen der vier Teilstudien der vorliegenden Dissertation.

3.3.1 Studie 1

Das Ziel der ersten Studie ist, das theoretisch abgeleitete und bisher noch nicht empirisch überprüfte kompetenzielle Feedback auf seine hypothetisch angenommenen positiven Effekte im Lernprozess zu untersuchen. Da es sich um eine erste Überprüfung der kompetenziellen Feedbackeffekte handelt, wird eine Vergleichsgruppe ohne Feedback gewählt; diese Baseline ermöglicht eine gezielte Aussage über die potenziell förderlichen oder evtl. hinderlichen (z.B. durch *cognitive overload*, van Merriënboer & Sweller, 2005) Effekte kompetenziellen Feedbacks.

Die Forschungsfrage der ersten Studie lautet deshalb: Welche Effekte zeigt kompetenzielles Feedback im Vergleich zu keinem Feedback auf die Performanz und Motivation von Lernenden?

Da aufgrund von Studie 1 keine Aussage getroffen werden kann, inwiefern kompetenzielles Feedback anderen Feedbackarten über- bzw. unterlegen ist, knüpft Studie 2 genau an dieser Fragestellung an.

3.3.2 Studie 2

Abgesehen von der Replikation der in Studie 1 gefundenen Effekte ist das Ziel der zweiten Studie der Vergleich der Lern- und Motivationsförderlichkeit kompetenziellen Feedbacks im Vergleich zu im schulischen Kontext gängigem Feedback, dem sogenannten sozialen Feedback (d.h. Einstufung der Schülerleistungen in Relation zum Klassendurchschnitt mithilfe von Noten). Zudem werden nicht nur Effekte auf Performanz und Motivation, sondern auch auf die Genauigkeit der Selbsteinschätzung (*calibration accuracy*) untersucht. Die Genauigkeit der Selbsteinschätzung wird als abhängige Variable aufgenommen, da vermutet wird, dass der dem Feedback (kompetenziell versus sozial) zugrundeliegende Vergleichsmaßstab (Kompetenzstufen versus Klassendurchschnitt) diese

Variable in unterschiedlichem Maße beeinflusst. Der im kompetenziellen Feedback angewandte Vergleichsmaßstab, das Verdeutlichen von erreichten und (noch) nicht erreichten Kompetenzstufen, enthält Informationen über das Lernziel, die erreichte Performanz in Relation zu diesem Lernziel wie auch zusätzliche Informationen zum verbesserten Erreichen des Lernziels bei zukünftigen Aufgabenbearbeitungen. Folglich kann angenommen werden, dass kompetenzielles Feedback zu einer genaueren Selbsteinschätzung führen sollte als das Wissen um die Einordnung der eigenen Leistung innerhalb der Klasse (soziales Feedback). Zudem wird in der zweiten Studie analysiert, ob dieser vermutete Unterschied in der Genauigkeit der Selbsteinschätzungen dem lernförderlichen Effekt kompetenziellen Feedbacks zugrundeliegt.

Die Forschungsfragen lauten demnach erstens: Welche Effekte des kompetenziellen Feedbacks auf Performanz, Motivation (Kompetenzerleben) und Genauigkeit der Selbsteinschätzung zeigen sich im Vergleich zu einem sozialen Feedback und einer Kontrollgruppe ohne Feedback? Und zweitens: Vermittelt die Genauigkeit der Selbsteinschätzungen den angenommenen positiven Effekt kompetenziellen Feedbacks auf die Performanz von Lernenden?

3.3.3 Studie 3

Das Ziel der dritten Studie ist die Erweiterung der in Studie 1 und 2 untersuchten (und als effektiv erwiesenen) Effekte kompetenziellen Fremdfedbacks um den Vergleich mit kompetenziellem Selbstfeedback und einer Kontrollgruppe ohne Feedback. Somit wird kompetenzielles Feedback in dieser Studie nicht mit anderen Feedbackarten wie beispielsweise sozialem Feedback verglichen, sondern der Effekt der Feedbackquelle (Fremdbewertung versus Selbstbewertung) wird manipuliert. In beiden Feedbackbedingungen ist das Feedback kompetenziell, wobei die Performanz entweder fremdbewertet

in Relation zu den Kompetenzstufen eingeschätzt und rückgemeldet wird (externes Feedback) oder die Lernenden selbst ihre Performanz in Relation zu den Kompetenzstufen bewerten (internes Feedback). Der Literatur zu vergleichenden Untersuchungen zwischen Fremd- und Selbstbewertung wird mit der dritten Studie eine systematische Analyse der Effekte eines als effektiv erwiesenen und für Fremd- und Selbstbewertung identischen Vergleichsmaßstabs (Kompetenzstufen) hinzugefügt. Effekte auf Performanz wie auch auf Genauigkeit der Selbsteinschätzung werden betrachtet.

Die Forschungsfragen lauten dabei erstens: Zeigen sich in der Selbstbewertung kompetenziellen Feedbacks die bereits für die kompetenzielle Fremdbewertung nachgewiesenen positiven Effekte auf Performanz und Genauigkeit der Selbsteinschätzungen im Vergleich zu keinem Feedback? Und zweitens: Ist der erwartete Unterschied in den Versuchsbedingungen durch die Genauigkeit der Selbsteinschätzungen begründet?

3.3.4 Studie 4

Die vierte Studie geht den in Studie 3 identifizierten differenziellen Effekten kompetenziellen Fremd- und Selbstfeedbacks auf Performanz nach. Neben der Replikation dieser Befunde steht im Zentrum, eine mögliche Ursache der differenziellen Effekte zu untersuchen. Als mögliche Ursache wird – anknüpfend an Peer-Feedbackstudien – die wahrgenommene Feedbackeffektivität in Betracht gezogen.

Dabei stellen sich folgende Forschungsfragen: (1) Nehmen Lernende mit Fremdfeedback (extern) das erhaltene Feedback als effektiver wahr im Vergleich zu Lernenden, die sich ihr Feedback selbst generieren (intern)? (2) Erklärt diese Divergenz in der wahrgenommenen Feedbackeffektivität die differenziellen Effekte externen versus internen kompetenziellen Feedbacks auf die Performanz?

4 LITERATUR

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 159-181. doi:10.1080/02602930600801928
- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27, 3-13. doi: 10.1111/j.1745-3992.2008.00118.x
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359-372. doi:10.1037/h0043445
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238. doi:10.3102/00346543061002213
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, 18, 5-25. doi: 10.1080/0969594X.2010.513678
- Bernholt, S. (2010). *Kompetenzmodellierung in der Chemie - Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*. Berlin: Logos Verlag.
- Bernholt, S., Neumann, K., & Nentwig, P. (Eds.) (2012), *Making it tangible - Learning outcomes in science education*. Münster: Waxmann.
- Bernholt, S., Parchmann, I., & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 217-243.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74. doi:10.1080/0969595980050102

- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529-549. doi:10.1007/BF00138746
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*, 399-413. doi:10.1080/02602930600679050
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research, 65*, 245-281. doi:10.3102/00346543065003245
- Bybee, R. (1997). *Achieving scientific literacy. From purposes to practices*. Portsmouth, NH: Heinemann.
- Connell, M. W., Sheridan, K., & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126-155). Cambridge, MA: Cambridge University Press.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*, 331-350. doi:10.1080/03075079912331379935
- Eggert, S., & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz. Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften, 12*, 177-194.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430. doi:10.3102/00346543059004395
- Fischer, H. E., & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht. *Der mathematische und naturwissenschaftliche Unterricht, 54*, 388-393.

- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304-315. doi:10.1016/j.learninstruc.2009.08.007
- Giere, R., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. Belmont, CA: Thomson Wadsworth.
- Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*, 21-34. doi:10.1080/00220670309596625
- Grube, C., Möller, A., & Mayer, J. (2007). Dimensionen eines Kompetenzstrukturmodells zum Experimentieren. In H. Bayrhuber, U. Harms, D. Krüger, A. Sandmann, A. Unterbrunner, A. Upmeyer zu Belzen, & H. Vogt (Eds.), *Ausbildung und Professionalisierung von Lehrkräften* (S. 31-34). Kassel: Universität Kassel.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht, 57*, 196-203.
- Harks, B., Rakoczy, K., Hattie, J., Klieme E., & Besser, M. (2011). *Indirect feedback effects on mathematics achievement*. Manuskript eingereicht zur Publikation.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi:10.3102/003465430298487
- Hattie, J., & Wollenschläger, M. (2012). *A conceptualization of feedback*. Manuskript eingereicht zur Publikation.
- Jäger, R. S. (2008). Leistungsbeurteilung. In W. Schneider, & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (S. 324-336). Göttingen: Hogrefe.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. Studien zum Physik- und Chemielernen*. Berlin: Logos Verlag.
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den naturwissenschaftlichen Fächern der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135-153.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: Bundesministerium für Bildung und Forschung.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. *Zeitschrift für Pädagogik*, 56, 64-74.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3-22). Göttingen: Hogrefe.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik*, 52, 876-903.

- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284. doi:10.1037//0033-2909.119.2.254
- Köller, O. (2005a). Bezugsnormorientierung von Lehrkräften: Konzeptuelle Grundlagen, empirische Befunde und Ratschläge für praktisches Handeln. In R. Vollmeyer, & J. C. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 189-202). Stuttgart: Kohlhammer.
- Köller, O. (2005b). Neue Besen kehren gut: Das Institut für Qualitätsentwicklung im Bildungswesen soll die nationalen Bildungsstandards in Deutschland überprüfen. *Psychologie in Erziehung und Unterricht*, *52*, 281-286.
- Köller, O., & Parchmann, I. (2012). Competencies: The German Notion of Learning Outcomes. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Making it tangible - Learning outcomes in science education* (pp. 165-185). Münster: Waxmann.
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Journal of Psychology*, *216*, 61-73. doi:10.1027/0044-3409.216.2.61
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*, 211-232. doi:10.3102/00346543047002211
- Kulhavy, R. W., & Stock, W. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*, 279-308. doi:10.1007/BF01320096
- Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. Dempsey & G. Ales (Eds.), *Interactive instruction and feedback* (pp. 3-20). Englewood Cliffs, NJ: Educational Technology Publications.

- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173-194. doi:10.1007/s11409-010-9056-2
- Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, 35, 135-156. doi:10.1080/02602930802687737
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger, & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung* (S. 177-186). Berlin: Springer.
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms, & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 63-79). Innsbruck: Studienverlag.
- Mischo, C., & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung. *Zeitschrift für Pädagogische Psychologie*, 9, 139-152.
- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.
- Narciss, S. (2006). *Informatives tutorielles Feedback*. Münster: Waxmann.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–143). Mahwah, NJ: Erlbaum.

- Nulty, D. D. (2011). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education*, 36, 493-507. doi:10.1080/02602930903540983
- Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction. Grading and feedback in mathematics classrooms. *Journal of Psychology*, 216, 110-123. doi:10.1027/0044-3409.216.2.111
- Rheinberg, R. (1980). *Leistungsbewertung und Lernmotivation*. Hogrefe: Göttingen
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider, & M. Hasselhorn (Eds.), *Handbuch Pädagogische Psychologie* (S. 178-186). Göttingen: Hogrefe.
- Rheinberg, F., Duschka, R., & Michels, U. (1980). Zielsetzung und Kausalattribution in Abhängigkeit vom Leistungsvergleich. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 12, 177-189.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. doi:10.1007/BF00117714
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5, 77-85. doi:10.1080/0969595980050104
- Schecker, H. (2012). Standards, competencies and outcomes. A critical view. In S. Bernholt, K. Neumann, & P. Nentwig (Hrsg.), *Making it tangible - Learning outcomes in science education* (S. 237-253). Münster: Waxmann.
- Schecker, H., & Parchmann, I. (2006): Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45-66.
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, 18, 455-463. doi:10.1006/ceps.1993.1034

- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189. doi:10.3102/0034654307313795
- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20, 291-303. doi:10.1016/j.learninstruc.2009.08.008
- Thorndike, E. L. (1913). *Educational psychology. Volume I: The original nature of man*. New York: Columbia University.
- Thorndike, R. L., & Hagen, E. P. (1971). *Cognitive Abilities Test*. Boston: Houghton Mifflin.
- Towler, L., & Broadfoot, P. (1992). Self-assessment in primary school. *Educational Review*, 44, 137-151. doi:10.1080/0013191920440203
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147-177.
- Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Zeitschrift für Pädagogik*, 56, 92-103.
- Vollmeyer, R., & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System. *Zeitschrift für Pädagogische Psychologie*, 12, 11-23.
- Vollmeyer R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15, 589-602. doi:10.1016/j.learninstruc.2005.08.001
- Waddington, D., Nentwig, P., & Schanze, S. (Eds.) (2007). *Standards in science education*. Münster: Waxmann.

- Walpuski, M., Kampa, N., Kauertz, A., & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht*, *61*, 323-326.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe & Huber.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*, 18-31. doi:10.1207/s1532690xci1601_2
- Zilker, I., Holliday, G. M., Fischer, H. E., Kauertz, A., Lederman, J. S., & Lederman, N. G., (2010). Are historical contexts suitable for assessing students' competences in the field of nature of science and scientific inquiry? In M.F. Taşar, & G. Çakmakçı (Eds.), *Contemporary science education research: international perspectives* (pp. 237-244). Ankara, Turkey: Pegem Akademi.

STUDIE 1

EFFEKTE KOMPETENZIELLER RÜCKMELDUNG BEIM WISSENSCHAFTLICHEN DENKEN

Wollenschläger, M., Möller, J., & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift für Pädagogische Psychologie*, 25, 197-202.

doi: 10.1024/1010-0652/a000040

ZUSAMMENFASSUNG

Kompetenzielle Rückmeldung¹ wurde in dieser Studie als Leistungsbewertung in Relation zu definierten Kompetenzstufen eines domänenspezifischen Kompetenzmodells verstanden. In einem Pre-Posttest-Design ($N = 38$) wurde eine Experimentalbedingung mit kompetenzieller Rückmeldung mit einer Kontrollgruppe ohne Rückmeldung verglichen. Die Ergebnisse zeigen, dass Lernende, die während der Bearbeitung von Aufgaben zum wissenschaftlichen Denken eine kompetenzielle Rückmeldung erhielten, besser abschnitten als Lernende, die keine Rückmeldung erhielten. Zudem zeigten sich tendenziell positive Effekte der kompetenziellen Rückmeldung auf motivationale Variablen, wie beispielsweise Interesse und Kompetenzerleben der Lernenden.

¹ In der Veröffentlichung von Studie 1 wurde der Begriff „Rückmeldung“ synonym für den Feedbackbegriff (siehe 1.1) verwendet.

ABSTRACT

In this study, competential feedback is defined as the evaluation of students' achievements in relation to defined competency levels of a domain-specific competence model. In a pre-post experimental design ($N = 38$) an experimental condition with competential feedback was compared to a control condition without any feedback. Students in the experimental condition showed a significantly better task performance in planning scientific experiments than students in the control condition. In addition, competential feedback tends to moderately increase certain motivational variables such as interest and perceived feeling of competence.

1 THEORETISCHER HINTERGRUND

Im Kontext der zunehmenden Fokussierung auf Bildungsziele und deren Konkretisierung in den Bildungsstandards hat die Entwicklung von Kompetenzmodellen in der pädagogisch-psychologischen und fachdidaktischen Forschung einen zentralen Platz eingenommen (Koeppen, Hartig, Klieme, & Leutner, 2008; Viering, Fischer, & Neumann, 2010). Für die pädagogische Praxis stellt sich die Frage, wie Kompetenzmodelle sowie Informationen aus kompetenzorientierter Diagnostik im Unterricht konkret umgesetzt und genutzt werden können (Klieme & Leutner, 2006). In dieser Studie steht die Forschungsfrage im Zentrum, inwiefern die Kompetenz und die Motivation der Lernenden durch eine kompetenzstufenorientierte Leistungsrückmeldung beeinflusst werden. Der theoretische Hintergrund dieser Studie verknüpft somit zwei Forschungsrichtungen: die Forschung zu Effekten von Rückmeldung im Lernprozess und die Kompetenzforschung. In einer ersten empirischen Überprüfung soll geklärt werden, ob die Rückmeldung der eigenen Leistung bezogen auf ein Kompetenzstufenmodell positive Effekte auf die Performanz und die Motivation von Lernenden hat im Vergleich zu einer Kontrollbedingung, in der das Kompetenzstufenmodell ebenfalls bekannt ist, die eigene Leistung aber nicht eingeschätzt wird. Diese Kontrollbedingung wurde erstens gewählt, da im Vergleich mit der Experimentalbedingung gezielt der Effekt der Rückmeldung identifiziert werden kann. Zweitens wurde sie gewählt, um einen empirischen Nachweis ihrer Effektivität zu erbringen; dies ist notwendig, da empirische Studien zeigen, dass nicht generell von einer positiven Wirkung von Feedback auszugehen ist (Kulhavy, 1977). Es könnte beispielsweise sein, dass die Lernenden aufgrund der hohen Informationsdichte und Komplexität der Rückmeldung gerade in einem Kompetenzstufenmodell mittels eines *cognitive overload* (van Merriënboer & Sweller, 2005) beeinträchtigt werden.

Im Folgenden werden zunächst die Effekte von Rückmeldungen im Hinblick auf das Ziel der Studie näher beschrieben, bevor anschließend die kompetenzielle Bezugsnorm theoretisch abgeleitet wird.

1.1 Effekte von Rückmeldungen

Rückmeldung wird definiert als Information, die einer Person in Bezug auf ihr eigenes Handeln und während oder nach einem Prozess oder Prozessschritt gegeben wird, um regulierend auf Prozess- oder Prozessschritt-Ebene einzuwirken (Narciss, 2006). Rückmeldung wird folglich durch die Stimuli definiert, die kontingent während oder nach der Bearbeitung von Aufgaben von einer externen Informationsquelle dargeboten werden (Hattie & Timperley, 2007). Somit entsprechen Rückmeldungen einer *post-response information* im lernpsychologischen Sinn, dementsprechend werden unterschiedliche Wirkmechanismen von Rückmeldung angenommen: Rückmeldung kann positive Effekte auf Performanz und Motivation haben, da sie als Bestätigung korrekter Antworten, als Korrektur falscher Antworten oder als Anreiz dienen kann, ein bestimmtes Ergebnis zu erreichen (Kulhavy & Wager, 1993; Narciss, 2006). Generell wird angenommen, dass Rückmeldung zu einer Anpassung von Verhalten an ein vorgegebenes Ziel führt (Kulhavy, 1977). Meta-Analysen identifizieren Rückmeldung als essenzielle Einflussgröße effektiver Lehr-/Lernprozesse (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996).

Zusammenfassend zeigen empirische Befunde eine positive Auswirkung von Rückmeldung oder Feedback auf die Performanz von Lernenden, (1) wenn die Rückmeldung eine Lücke zwischen Ist- und Sollzustand verdeutlicht und dadurch Unsicherheit verringert, (2) wenn durch die Rückmeldung die kognitive Belastung reduziert wird sowie (3) wenn ungünstige Strategien bei der Aufgabenlösung verändert werden (Bangert-Drowns et al., 1991; Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996, Mory,

2004; Shute, 2008). Zudem ist Rückmeldung umso effektiver, je aufgabenbezogener, spezifischer und detaillierter sie formuliert ist; dabei sollten Informationen vor allem hinsichtlich der Frage enthalten sein, wie die Aufgabenbearbeitung zukünftig verbessert werden kann (Shute, 2008). In Studien, welche die Auswirkung von Rückmeldungen auf die Motivation oder das Selbstkonzept der Lernenden untersuchten (Vollmeyer & Rheinberg, 2005), zeigten sich ebenfalls positive Effekte einer Rückmeldung, die Information über die weitere Aufgabenbearbeitung bereit hielt (Rakoczy, Klieme, Bürgermeister & Harks, 2008).

Die vorliegende Studie hat sich auf spezifische Annahmen bezüglich Kompetenzerleben und Interesse der Lernenden fokussiert, da diese in engem Zusammenhang mit dem Prozess der Selbstbewertung (Bandura, 1977; Deci & Ryan, 1980) und somit der Rückmeldewirkung stehen. In Erwartungs-Wert-Theorien (z.B. Atkinson, 1957; Wigfield & Eccles, 2000) wird deutlich, dass Motivation zum einen von Kompetenzeinschätzungen, zum anderen vom Wert, welcher der Tätigkeit zugesprochen wird, abhängt. Zudem wird angenommen, dass Kompetenzeinschätzungen, z.B. aufgrund von Rückmeldeprozessen der eigenen Leistung, zentral für die Interessensgenese sind (z.B. Wigfield & Eccles, 2000). Das *cognitive-motivational model* (Vollmeyer & Rheinberg, 1998) beinhaltet die Annahme, dass die Rückmeldung, die den Lernenden aufzeigt, wie gut sie bei einer Aufgabe abgeschnitten haben, motivierend wirkt und dazu führt, sich bei zukünftigen Aufgabenbearbeitungen vermehrt anzustrengen (Vollmeyer & Rheinberg, 2005). Folglich kann angenommen werden, dass die kompetenzielle Rückmeldung motivierend wirkt, da sie so formuliert ist, dass sie aufzeigt, welche Kompetenzstufe schon erreicht wurde. Dies sollte zu einer erhöhten Kompetenzeinschätzung führen und in der Folge zu einem erhöhten Interesse der Lernenden. Zudem zeigten Rückmeldungen, die auf Basis unterschiedlicher Bezugsnormen gestaltet wurden, differenzielle Effekte auf die Motivation der Lernenden (Rheinberg, 2008; Rheinberg, Duscha, & Michels, 1980). Bezugsnormen beschreiben

unterschiedliche Kriterien, nach denen Leistung bewertet wird. Im schulischen Kontext können sich Leistungsrückmeldungen an sozialen, individuellen oder kriterialen Bezugsnormen orientieren (Rheinberg, 2008). Die soziale Bezugsnorm beinhaltet den Leistungsvergleich eines Lernenden mit seiner sozialen Bezugsgruppe; die individuelle Bezugsnorm umfasst den Vergleich mit einem vorhergehenden Leistungsergebnis desselben Lernenden; die kriteriale Bezugsnorm vergleicht die erreichte Leistung des Lernenden mit einem Standard, wie beispielsweise dem Erwartungshorizont der Lehrkraft bezüglich der gestellten Aufgabe (Rheinberg, 2008). Zusammenfassend kann gesagt werden, dass die individuelle sowie die kriteriale Bezugsnorm motivationsfördernder sind als die soziale Bezugsnorm (Rheinberg, 2008; Rheinberg et al., 1980).

1.2 Kompetenzielle Bezugsnorm

Basierend auf der kriterialen Bezugsnorm und aufgrund der Kompetenzorientierung unterrichtlichen Handelns kann nach Jäger (2008) eine sogenannte fähigkeitsorientierte Bezugsnorm abgeleitet werden; dabei sollen Aussagen über die Fähigkeit in Relation zu definierten Kompetenzstufen gegeben werden. Im Unterschied zur kriterialen Rückmeldung sollen dabei die Kompetenzstufen der fähigkeitsorientierten Bezugsnorm aus Kompetenzmodellen abgeleitet und nicht wie bei der kriterialen Bezugsnorm subjektiv normativ gesetzt werden.

Solche Kompetenzmodelle stellen die Basis einer validen Kompetenzmessung dar, mit deren Hilfe die erhobenen Messdaten in Bezug auf ein zugrundeliegendes theoretisches Modell interpretiert werden können (Koeppen et al., 2008). Kompetenzmessung ermöglicht dabei eine summative wie auch formative Evaluation des Kompetenzstandes; und dies sowohl auf Systemebene, wenn beispielsweise Bildungsprogramme evaluiert werden, als auch auf individueller Schülerebene. Auf individueller Ebene bietet die Kompetenzmessung

die Möglichkeit, Personen Kompetenzstufen zuzuordnen, d.h. Aussagen über den Leistungsstand eines Lernenden bezüglich einer spezifischen Kompetenz zu treffen.

2 ENTWICKLUNG DER FRAGESTELLUNG

Die Ergebnisse der Meta-Analysen zu den Effekten von Rückmeldungen machen deutlich, dass eine effektive Rückmeldung konkret, spezifisch und bezogen auf eine Aufgabe formuliert ist. Des Weiteren sollte sie einen objektiven Vergleichsstandard beinhalten und so einen Vergleich der aktuellen Performanz mit dem erwünschten Lernziel ermöglichen. Diese Studie zielt darauf ab, Effekte einer Bezugsnorm zu untersuchen, die sich auf ein Kompetenzstufenmodell zum naturwissenschaftlichen Unterricht (Planung von Experimenten, siehe Mayer, Grube, & Möller, 2008) bezieht. Die Leistungsrückmeldung geschieht in dieser Studie in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells. Da die in dieser kompetenziell genannten Bezugsnorm enthaltenen Informationen die genannten Kriterien günstiger Rückmeldungen erfüllen und eine Erweiterung einer Rückmeldung basierend auf der kriterialen Bezugsnorm darstellen, sollte die kompetenzielle Rückmeldung einen positiven Effekt auf die Performanz und die Motivation von Lernenden haben.

3 METHODE

In einer experimentellen Studie wurden Effekte einer kompetenziellen Rückmeldung auf die Performanz und die Motivation von Lernenden untersucht. Während in der Experimentalgruppe den Probanden mitgeteilt wurde, auf welcher Kompetenzstufe sie sich befinden, erhielt die Kontrollgruppe keine solche Rückmeldung. Insgesamt ergibt sich ein zweifaktorielles 2 (*Rückmeldung*: mit vs. ohne) x 5 (*Zeitpunkt*: t1 vs. t2 vs. t3 vs. t4 vs. t5)-Design mit Messwiederholung auf der zweiten Variable.

3.1 Stichprobe

Die Probanden waren Schülerinnen und Schüler aus vier 8. Klassen an Kieler Gymnasien. Die Teilnahme erfolgte freiwillig. Die Analytestichprobe bestand aus $N = 38$ Schülerinnen und Schülern ($M = 13.24$ Jahre, $SD = 0.54$, 21 weiblich).

3.2 Variablen

UV 1: Rückmeldung. Als Stufen dieser unabhängigen Variable fungierten eine Bedingung mit kompetenzieller Rückmeldung und eine Bedingung ohne kompetenzielle Rückmeldung. Die Probanden wurden dieser Bedingung per Zufall (Losverfahren) zugewiesen. Basierend auf dem Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008) wurde für die Teilkompetenz „Experimente planen“ ein Rückmeldebogen erstellt. Die Niveaustufen des Kompetenzmodells sind nach zunehmender Komplexität graduiert und reichen von der „Untersuchung eines Faktors“, „Untersuchung von Zusammenhängen“, „Kontrollierte Untersuchung auf Basis von Konzeptverständnis“, „Elaborierte Untersuchung allgemeiner Zusammenhänge“ bis hin zum „Selbständigen Lösen von offenen Problemen“. Für die Teilkompetenz „Experimente planen“ lauten die fünf Niveaustufen (1) „Eine Variable identifizieren“, (2) „Veränderte und zu messende Variable in Beziehung setzen“, (3) „Kontrollvariablen berücksichtigen“, (4) „Stichprobe, Messwiederholung und Versuchsdauer berücksichtigen“, (5) „Untersuchungsmethoden, Genauigkeit und Fehler abwägen“. Die im Modell beschriebenen Kompetenzstufen zum „Experimente planen“ wurden schülergerecht umformuliert, sodass die jeweils erreichte Kompetenzstufe (5-Stufen-Skala) auf dem Rückmeldebogen angekreuzt werden konnte („Bei der Planung eines Experiments hast du...: 1) ...gesagt, was du untersuchen willst. Du hast eine Variable genannt. Eine Variable ist die zu untersuchende Bedingung, die im Experiment verändert, d.h. variiert wird., 2) ...diese

Variable verändert und gesagt, wie du die Veränderungen messen willst., 3) ...andere Einflussfaktoren, die du nicht untersuchen willst, gleich gehalten., 4) ...gesagt, an wie vielen Untersuchungsgegenständen, wie oft oder wie lange du das Experiment durchführen willst., 5) ...über mögliche Fehler oder Schwächen in deiner Planung nachgedacht und hast dann eine neue Untersuchung entworfen.“). Die Instruktion auf dem Rückmeldebogen erläuterte, dass das Kreuz die erreichte Kompetenzstufe bei der letzten Aufgabenbearbeitung verdeutlicht und die nicht angekreuzten Felder aufzeigen, wie die nächste Aufgabenbearbeitung verbessert werden kann. In der Kontrollbedingung erhielten die Probanden keine Einstufung der eigenen Performanz, jedoch waren die Informationen des Kompetenzmodells bekannt. Alle Probanden erhielten eine standardisierte Einführung in das wissenschaftliche Denken, die inhaltlich den Kompetenzstufen der Teilkompetenz „Experimente planen“ aus dem Kompetenzmodell von Mayer et al. (2008) entsprach. Somit sollte gewährleistet werden, dass sich die Versuchsbedingungen nicht hinsichtlich der Informationen über das Kompetenzmodell unterschieden.

UV 2: Zeitpunkte. Die Probanden bearbeiteten insgesamt fünf Aufgaben zum Experimentieren (Mayer et al., 2008). Zu den Zeitpunkten t1 bis t5 wurden daher ihre Leistungen bewertet.

AV 1: Performanz. Die erste abhängige Variable war die Performanz beim wissenschaftlichen Denken im Sinne der Güte der Bearbeitung von Aufgaben zur Planung von Experimenten. Die Performanz wurde mithilfe von fünf offenen Paper-Pencil-Aufgaben (Mayer et al., 2008) erhoben. Basierend auf den raschskalierten Schwierigkeitsparametern aus der Studie von Mayer et al. (2008) wurden sie nach zunehmender Aufgabenschwierigkeit angeordnet. In den Aufgaben werden wissenschaftliche Fragestellungen in alltagsnahen Situationen beschrieben, woraufhin die Lernenden jeweils ein Experiment planen sollen, welches die beschriebene Fragestellung beantwortet (z.B. „Tom möchte seiner Mutter zum

Geburtstag Rosen schenken. In einer Gartenzeitschrift hat er gelesen, dass Rosen besser in Blumenerde wachsen als in Gartenerde. Bevor er seiner Mutter die Rosen schenkt, möchte er diese Aussage aus der Gartenzeitschrift prüfen. Plane ein Experiment, mit dem diese Aussage naturwissenschaftlich überprüft werden kann. Beschreibe das Experiment möglichst genau.“). Die offenen Antworten zu den Aufgaben wurden nach der Versuchsdurchführung erneut anhand des bestehenden Kodiermanuals von zwei hypothesen- und bedingungsblinden Ratern auf den Kompetenzstufen von 1-5 kodiert. Die zufallskorrigierte Interrater-Reliabilität der Aufgaben 1-5 reichte von $\kappa = 0.74$ bis $\kappa = 0.96$ (Cohen, 1960) und war somit zufriedenstellend (Landis & Koch, 1977). Die Übereinstimmung dieser Ratings mit den Aufgabenbewertungen, die während der Versuchsdurchführung kodiert wurden, lag zwischen $\kappa = 0.67$ bis $\kappa = 0.85$. Da die Ratings während der Versuchsdurchführung nicht bedingungsblind waren, gingen in die folgende statistische Hypothesenprüfung die Rohwerte ein, auf die sich beide hypothesenblinden Rater geeinigt hatten.

AV 2: Motivation. Als zweite abhängige Variable wurde die Motivation der Lernenden bezüglich dieser Aufgaben mit der Kurzskala intrinsischer Motivation (KIM) (Wilde, Bätz, Kovaleva & Urhahne, 2009) untersucht. Hinsichtlich der Variable Motivation liegt ein ausschließlich einfaktorielles Design vor, da Motivation nach der Bearbeitung der fünf Aufgaben erfasst wurde. Dieses Vorgehen ist der ökologischen Validität geschuldet; es sollte eine natürliche Situation der Aufgabenbearbeitung ermöglicht werden. Die KIM besteht aus vier Subskalen mit jeweils drei Items, die an die Domäne angepasst wurden: „Interesse“ (z.B. „Ich fand das Planen von Experimenten sehr interessant“), „wahrgenommene Kompetenz“ (z.B. „Mit meiner Leistung bei der Planung von Experimenten bin ich zufrieden“), „wahrgenommene Wahlfreiheit“ (z.B. „Beim Planen von Experimenten konnte ich so vorgehen, wie ich es wollte“) und „Anspannung“ (z.B. „Beim Planen von Experimenten fühlte ich mich angespannt“). Auf einer Ratingskala von

1 = „stimmt nicht“ bis 4 = „stimmt genau“ sollten die Lernenden angeben, wie sehr sie den Aussagen zustimmten. Die Reliabilität der Subskalen der KIM „Interesse“ ($\alpha = .94$), „wahrgenommene Kompetenz“ ($\alpha = .85$) und „wahrgenommene Wahlfreiheit“ ($\alpha = .81$) war zufriedenstellend, für „Anspannung“ ($\alpha = .58$) galt dies nicht.

3.3 Versuchsablauf

Das Experiment wurde in Gruppen mit jeweils bis zu zehn Lernenden durchgeführt. Die Dauer betrug ca. 60 Minuten. Die Lernenden erhielten zunächst einen Text mit einer standardisierten Einführung in das wissenschaftliche Denken. Anschließend bearbeiteten sie die fünf Aufgaben zur Planung von Experimenten. Es war keine Zeitbegrenzung vorgegeben. In der Experimentalbedingung kodierte die Versuchsleiterin die Antworten der offenen Aufgaben anhand eines bestehenden Kodiermanuals (Mayer et al., 2008), wies den Probanden damit eine Kompetenzstufe zu und füllte den Rückmeldebogen entsprechend aus (siehe Abschnitt 3.2). Dieser wurde an die Lernenden zurückgegeben, bevor sie mit der nächsten Aufgabe fortfuhren. Zum Schluss folgte ein Fragebogen, der die KIM wie auch demographische Items beinhaltete.

4 ERGEBNISSE

Es wurden eine Varianzanalyse mit Messwiederholung für die Performanz und eine multivariate Varianzanalyse für die Motivationsskalen gerechnet. Die deskriptiven Werte sind in Tabelle 1 dargestellt.

Tabelle 1

Mittelwerte (Standardabweichungen) und Effektstärken der Aufgaben 1-5 zur Planung eines Experiments und der Subskalen der Kurzskala intrinsischer Motivation (KIM).

	Rückmeldung		
	kompetenziell ($n = 19$)	ohne ($n = 19$)	d
Aufgabe 1	3.26 (0.87)	3.63 (0.76)	0.45
Aufgabe 2	4.32 (0.67)	3.95 (0.62)	0.57
Aufgabe 3	4.21 (0.92)	3.42 (1.07)	0.79
Aufgabe 4	4.26 (1.19)	3.11 (1.29)	0.93
Aufgabe 5	3.89 (1.20)	2.63 (0.96)	1.16
Interesse	3.25 (0.67)	2.79 (0.90)	0.48
Wahrgenommene Kompetenz	3.05 (0.65)	2.65 (0.79)	0.55
Wahrgenommene Wahlfreiheit	3.40 (0.49)	3.19 (0.72)	0.34
Anspannung	2.28 (0.59)	2.38 (0.64)	0.16

Anmerkungen: Die Performanz konnte einen Wert von 1-5 annehmen; ein höherer Wert bedeutet eine höhere Performanz. Die KIM beinhaltet eine vierstufige Ratingskala von 1 = „stimmt nicht“ bis 4 = „stimmt genau“.

Performanz. In einer zweifaktoriellen Varianzanalyse mit den Faktoren *Rückmeldung* (kompetenziell vs. ohne)**Zeitpunkte* (t1 – t5) mit Messwiederholung auf dem Faktor *Zeitpunkte* zeigte sich ein signifikanter Haupteffekt für den Faktor *Rückmeldung* ($F(1,36) = 12.56$; $p < .01$; $\eta_p^2 = .26$). Lernende mit kompetenzieller Rückmeldung zeigten ab t2 insgesamt eine höhere Performanz (t2-t5: $M = 4.17$; $SD = 0.59$) als Lernende ohne kompetenzielle Rückmeldung (t2-t5: $M = 3.27$; $SD = 0.66$). Auch der Haupteffekt des Faktors *Zeitpunkte* wurde signifikant ($F(1,3.17) = 5.28$; $p < .01$; $\eta_p^2 = .13$). Hier fällt auf, dass die Performanz der Lernenden mit jeder Aufgabe abnahm; dies ist in der zunehmenden Schwierigkeit der Aufgaben begründet. Auch die Interaktion der Faktoren *Rückmeldung***Zeitpunkte* wurde signifikant ($F(1,3.17) = 5.19$; $p < .01$; $\eta_p^2 = .13$).

Angeschlossene paarweise Vergleiche zeigten zwischen Aufgabe 1 und Aufgabe 2 einen Zuwachs von t1 zu t2 nur in der Experimentalgruppe sowie eine stärkere Abnahme der Performanz in der Kontroll- als in der Experimentalbedingung in den Aufgaben 2 und 3 gegenüber Aufgabe 5.

Motivation. Es zeigte sich kein signifikanter Effekt der kompetenziellen Rückmeldung auf die Gesamtskala der KIM ($F(4,33) = 0.88; p = .48$). Analysiert man dennoch die einzelnen Skalen, ergeben sich tendenzielle Effekte in den Subskalen „Interesse“ ($F(1,36) = 3.11; p < .10; \eta_p^2 = .08$) und „Kompetenzerleben“ ($F(1,36) = 2.96; p < .10; \eta_p^2 = .08$). Wie in Tabelle 1 ersichtlich, waren das Interesse und das Kompetenzerleben in der Experimentalbedingung höher als in der Kontrollbedingung. In den Subskalen „wahrgenommene Wahlfreiheit“ ($F(1,36) = 1.10; p = .30$) und „Anspannung“ ($F(1,36) = 0.28; p = .60$) war der Unterschied zwischen den Versuchsbedingungen statistisch nicht bedeutsam.

5 DISKUSSION

Aus der Verknüpfung von Rückmeldeforschung und Kompetenzforschung wurde eine innovative theoriegeleitete Variante der Leistungsrückmeldung per kompetenzieller Bezugsnorm untersucht. Die kompetenzielle Bezugsnorm stellt eine Erweiterung der kriterialen Bezugsnorm dar, indem die von einer Person erreichten und noch zu erreichenden Stufen eines domänenspezifischen Kompetenzmodells dargestellt wurden. Das Ziel der Studie war eine erste empirische Klärung der Frage, ob die kompetenzielle Rückmeldung im Vergleich zu keiner Rückmeldung positive Effekte auf die Performanz und Motivation von Lernenden aufweist.

Hypothesenkonform waren positive Effekte der kompetenziellen Rückmeldung auf die Performanz feststellbar. Lernende mit kompetenzieller Rückmeldung schnitten nach

Erhalt der Rückmeldungen in allen Aufgaben besser ab als Lernende ohne kompetenzielle Rückmeldung. Einschränkend kann in der vorliegenden Studie allerdings nicht von einem Lernzuwachs durch die kompetenzielle Rückmeldung gesprochen werden, da die verwendeten Aufgaben unterschiedliche Schwierigkeiten beinhalteten. Trotzdem ist es gelungen zu zeigen, dass Schülerinnen und Schüler von der kompetenziellen Rückmeldung profitierten. Dieser Befund steht erstens in Einklang mit den Befunden der Meta-Analysen, die zeigen, dass Rückmeldungen dann eine Performanzsteigerung begünstigen, wenn sie spezifische Informationen über die weitere Aufgabenbearbeitung enthalten (Shute, 2008).

Zweitens scheint die Aufgabenbearbeitung nicht durch einen möglichen *cognitive overload* durch die Informationsdichte in der Rückmeldung beeinträchtigt worden zu sein.

Bezüglich der motivationalen Variablen hatte die kompetenzielle Rückmeldung nur schwach positive Auswirkungen auf das Interesse und das Kompetenzerleben der Lernenden. Auch dieser Befund spricht zumindest dafür, dass die kompetenzielle Rückmeldung bei der Aufgabenbearbeitung nicht als ablenkend oder überfordernd empfunden wurde. Zukünftige Studien sollten klären, ob Interesse- und Kompetenzskalen, die diese Konstrukte breiter als mit nur drei Items abbilden, Unterschiede zwischen kompetenziellen und anderen Rückmeldungen aufzeigen.

Allerdings müssen die Ergebnisse bezüglich folgender Konfundierung diskutiert werden: Die Probanden in der Experimentalbedingung erhielten im Rückmeldebogen Informationen über das Kompetenzmodell und Information über die Einstufung der eigenen Leistung im Kompetenzmodell. Durch die wiederholte Rückmeldung in den Zeitpunkten t1-t5 war somit in der Experimentalbedingung eine temporale Vergleichsinformation gegeben, die mit der kompetenziellen Rückmeldung konfundiert ist. Die Ergebnisse zeigen jedoch einen Zuwachs von t1 zu t2 nur in der Experimentalgruppe, also genau dann, wenn

noch keine solche Konfundierung vorliegt. Dies spricht gegen die Idee, dass die Rückmeldung nur deshalb wirkt, weil sie eine temporale Vergleichsinformation beinhaltet.

Zusammenfassend bietet die Studie eine erste empirische Evidenz für einen positiven Effekt einer aus Kompetenzmodellen theoretisch abgeleiteten Rückmeldung auf Performanz (und eingeschränkt auf Motivation). Dieser Befund bietet die empirische Basis für weitere Untersuchungen zu der Fragestellung, inwiefern kompetenzielle Rückmeldung anderen Rückmeldearten über- oder unterlegen ist. Dazu sollte die kompetenzielle Rückmeldung mit anderen Rückmeldungsbedingungen kontrastiert werden, beispielsweise mit sozial und längsschnittlich vergleichenden Rückmeldungen. Zudem sollten auch andere Kompetenzbereiche in den Blick genommen werden. Sofern diese vergleichenden Studien die positiven Befunde bestätigen, wird die kompetenzielle Rückmeldung für die Praxis relevant und nutzbar. Sie könnte dann im Experimentalunterricht oder in Schülerlaboren (Brandt, Möller, & Kohse-Höinghaus, 2008) eingesetzt werden. Bis dahin bleibt des Weiteren zu prüfen, welche zugrundeliegenden Prozesse die Leistungssteigerungen bedingen. Möglicherweise enthalten die Kompetenzstufenmodelle Informationen, aus denen Probanden günstige Bearbeitungsstrategien und Aufgabenlösungen ableiten können, wie dies etwa beim beispielbasierten Lernen (siehe Atkinson, Derry, Renkl, & Wortham, 2000; Mackensen-Friedrichs, 2009; Renkl, 1997, Schmidt-Weigand, Hänze, & Wodzinski, 2009) der Fall ist.

6 LITERATUR

- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181-214.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191-215.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213-238.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7-74.
- Brandt, A., Möller, J., & Kohse-Höinghaus, K. (2008). Was bewirken außerschulische Experimentierlabors? Ein Kontrollgruppenexperiment mit Follow up-Erhebung zu Effekten auf Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, *22*, 5-12.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.
- Deci, E. L., & Ryan, R. M. (1980). Self-determination theory: When mind mediates behavior. *Journal of Mind & Behavior*, *1*, 33-43.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81-112.
- Jäger, R. S. (2008). Leistungsbeurteilung. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 324-336). Göttingen: Hogrefe.

- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik*, 52, 876-903.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008): Current issues in competence modeling and assessment. *Journal of Psychology*, 216, 61-73.
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232.
- Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. V. Dempsey, & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 3-20). Englewood Cliffs: Educational Technology Publications.
- Landis, G., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Mackensen-Friedrichs, I. (2009). Die Rolle von Selbsterklärungen aufgrund vorwissensangepasster, domänenspezifischer Lernimpulse beim Lernen mit biologischen Beispielaufgaben. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 155-172.
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms, & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 63-79). Innsbruck: Studienverlag.

- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.
- Narciss, S. (2006). *Informatives tutorielles Feedback*. Münster: Waxmann.
- Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction. *Zeitschrift für Psychologie*, 216, 111-124.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21, 1-29.
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 178-186). Göttingen: Hogrefe.
- Rheinberg, F., Duscha, R., & Michels, U. (1980). Zielsetzung und Kausalattribution in Abhängigkeit vom Leistungsvergleich. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 12, 177-189.
- Schmidt-Weigand, F., Hänze, M., & Wodzinski, R. (2009). Komplexes Problemlösen und ausgearbeitete Lösungsbeispiele: Die Bedeutung von der Anregung strategischen Verhaltens und des Einblendens von Lösungsschritten, *Zeitschrift für Pädagogische Psychologie*, 23, 129-138.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147-177.
- Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Zeitschrift für Pädagogik*, 56, 92-103.

- Vollmeyer, R., & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System, *Zeitschrift für Pädagogische Psychologie*, 12, 11-23.
- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15, 589-602.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 31-45.

STUDIE 2

COMPETENTIAL FEEDBACK EFFECTS ON PERFORMANCE: ARE THEY MEDIATED BY CALIBRATION?

Wollenschläger, M., Hattie, J., Möller, J., & Harms, U. (2011). *Competential Feedback Effects on Performance: Are They Mediated by Calibration?*

Manuskript eingereicht zur Publikation.

ABSTRACT

Competential feedback is defined as the evaluation of students' achievements in relation to defined competency levels within a domain-specific competence model. This study aimed at investigating effects of this criterion-related feedback on performance, motivation and calibration. In a pre-post experimental design, randomly assigned secondary school students ($N = 120$) with competential feedback were compared to a social feedback condition (the students are given marks relative to their classmates' achievements), and to a control condition without any feedback at all. Students in the competential feedback condition showed a significantly better task performance in planning scientific experiments, perceived themselves as being more competent, and were also more accurate in their self-evaluative performance judgements when compared to the social and the control condition. This calibration accuracy mediated the positive effect of competential feedback on performance. The study indicates that competence models can provide a theoretically sound basis for providing worthwhile achievement feedback in schools.

1 INTRODUCTION

Feedback has been shown to have one of the most effective influences on learning outcomes, but these feedback effect sizes can vary considerably and not all feedback types lead to increases in learning (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie, 2009; Kluger & DeNisi, 1996). Therefore, further research needs to be undertaken to investigate more specifically how feedback works effectively in the classroom and learning process (Hattie, 2009). Shute (2008) argued that a worthwhile way to undertake such research is to concentrate on task specific feedback and mediational effects on the impact of feedback on performance. At the same time, feedback based on competence models has not yet been investigated sufficiently (Koeppen, Hartig, Klieme, & Leutner, 2008). Feedback based on competence models appears to affect learning outcomes when compared to other commonly used feedback methods (Wollenschläger, Möller, & Harms, 2011). This study combines these two research questions – understanding the way feedback works in classes, and the possible power of competence feedback.

1.1 Competence models

Competencies are defined as context-specific cognitive ability constructs which relate to situations and demands in domains concerning specific areas of learning (Klieme & Leutner, 2006). In the last decade, the term “competence” has been increasingly used in educational research and psychology (Connel, Sheridan, & Gardner, 2003; Klieme, Hartig, & Rauch, 2008; Klieme, Bürgermeister, Harks, Blum, Leiß, & Rakoczy, 2010; Weinert, 2001). Defining competencies allows differentiating between general, context-independent abilities and context-specific abilities. Unlike general abilities, competencies are acquired by learning and experience; thus they can be influenced by external interventions and educational processes (Hartig & Klieme, 2006). Competence models comprise the internal structure as

well as different levels of competencies with regard to domain-specific performance and take into account developmental processes (Koeppen et al., 2008). It is conjectured that feedback based on the assessed levels of a competence model can improve students' learning (Koeppen et al., 2008), especially when compared to a social feedback which is commonly used in classrooms (Wollenschläger et al., 2011).

1.2 Feedback effects

Feedback is defined as the information which is given concerning a task or a learning process (Hattie & Timperley, 2007; Narciss, 2008). Providing feedback in classrooms is found to have a very powerful impact on students' learning with an overall effect size of $d = 0.79$ (based on 12 meta-analyses, 196 studies and 6972 effect sizes; Hattie, 2009). But feedback effects are not necessarily beneficial and reported effect sizes can vary considerably (Bangert-Drowns et al., 1991; Hattie, 2009; Kulhavy, 1977). Therefore, meta-analyses have identified certain aspects making feedback effective on learning: It has been shown that feedback should focus on correct answers, needs to be task-specific, should elucidate specific goals telling learners the what, how and why of the task, and most importantly, should include cues on the current state of the learning process as well as on the how to improve task performance (Black & Wiliam, 1998; Hattie, 2009, 2012; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mory, 2004; Sadler, 1989, 1998; Shute, 2008). As summarized in Hattie and Timperley's feedback model (2007, p. 87), effective feedback has to answer one of the three questions: "Where am I going, How am I going, and Where to next". Beyond that, feedback needs to be adapted to the learner and to the task (Hattie & Gan, 2011). Thus, research needs to be done in order to evaluate which feedback works best for which tasks (Duijhouwer, Prins, & Stokking, in press; Hattie & Gan, 2011; Narciss & Huth, 2006; Rakoczy, Klieme, Bürgermeister, & Harks, 2008; Van der Kleij, Eggen, Timmers, & Veldkamp, 2012).

Feedback in classrooms can be based on social, individual or criterion-related reference norms (for an overview see Rheinberg & Krug, 1999 or related work Ames, 1992; Covington, 2001). Reference norms are external standards as to which students' achievements are evaluated: The individual student's performance is evaluated and fed back in relation to either other students' learning outcomes (social feedback), to learning outcomes which the same student has previously achieved (individual feedback), or to an external objective standard (criterion-related feedback) (Rheinberg, 2008). Feedback which is based on these different reference norms influences cognitive and motivational variables in academic learning (Krampen, 1987; Mischo & Rheinberg, 1995). To support cognitive as well as motivational learning outcomes, empirical results in self-regulation literature emphasize that feedback should be presented in a way which helps students to value their skill development (Cleary & Zimmerman, 2004) and to focus on enhancing standards of performance (Kitsantas & Zimmerman, 2006). This is considered when providing a competence model as a theoretically based criterion-related standard to evaluate students' achievements – such feedback is termed here competential feedback.

1.3 Mediation of feedback effects on performance

Feedback effects depend on students' motives and goals (Butler & Winne, 1995; Nolen, 1996). Hence, feedback studies need to take into account underlying variables mediating feedback influences on performance (Stone, 2000; Vollmeyer & Rheinberg, 2005). For example, mediational effects have been studied for process-oriented feedback showing an indirect feedback effect via perceived usefulness on intrinsic motivation and mathematics achievement (Harks, Rakoczy, Hattie, Klieme, & Besser, 2011).

This study focused on calibration accuracy as one potential mediator since feedback information is described as one critical factor leading to accurate calibration (Stone, 2000).

Calibration is defined as the magnitude of consent between learners' true and self-estimated test performance (Schraw, Potenza, & Nebelsick-Gullet, 1993). Well calibrated learners' test performance equals their perceived test performance. Divergence in perceived performance judgments displays either over- or underconfidence. Students are often overconfident about their test performance (Pajares & Miller, 1997; Schraw et al., 1993). Overconfidence was found to impede performance whereas accurate calibration leads to effective learning strategies and performance increase (Bol, Hacker, O'Shea, & Allen, 2005; Chen, 2002; Kruger & Dunning, 1999; Horgan, 1990; Labuhn, Zimmerman, & Hasselhorn, 2010; Schraw et al., 1993; Winne & Jamieson-Noel, 2002). In this study, it is examined whether competential feedback leads to more accurate performance judgments, thus leading to a better task performance.

1.4 The present study

This study investigates how a competence model can provide a theoretically sound basis for achievement feedback in schools. Students' achievements are evaluated in relation to defined competency levels of a domain-specific competence model which comprises an innovative but theoretically based extension of the criterion-related reference norm. This innovative feedback is called "competential" to clarify that students' achievements are evaluated and fed back relative to prescribed competence levels. So far, competential feedback has been investigated in comparison to a control group without any feedback information in the domain of scientific reasoning (Wollenschläger et al., 2011). The results from this study showed that competential feedback positively influenced students' performance as well as motivation when they planned scientific experiments.

The aim of this study is to extend these findings by comparing competential feedback not only to a control group without any feedback but also by comparing it to a commonly

used feedback in schools (marks) in the domain of scientific reasoning in science classes. Beyond that, an underlying variable (i.e. calibration accuracy) of the positive effects of competential feedback on performance will be investigated.

1.5 Hypotheses

A major focus of this study relates to the effects of competential feedback compared to social feedback (the students are given marks relative to their classmates' achievements), and to a control condition where no feedback is given at all. These feedback effects are investigated with respect to performance, motivation and calibration in the domain of scientific reasoning (i.e. planning experiments).

Competential feedback focuses on correct answers, is task-specific, goal-centered and informative as to how to improve task performance. Feedback which is based on competence models, illustrating reached and desired competence levels, therefore includes information about the actual as well as the potential performance elucidating a rubric of desired performance. In summary, this method addresses all aspects of effective feedback (where should the student go, how does the student go, and where does the student go to next). Thus, in the *first* hypothesis, it is assumed that competential feedback has a more beneficial influence on students' task performance than when the students are only informed on their achievement relative to their classmates' achievements, or when they are not given any feedback at all.

By elucidating goals and telling students their achievements relative to the goals, competential feedback enables students to value their skill development. Hence, competential feedback is assumed to beneficially influence motivational learning outcomes such as students' interest and perceived competence (*second* hypothesis), and calibration such as less biased and more accurate self performance judgments (*third* hypothesis) when compared to

feedback showing students' achievements relative to their classmates' achievements and to the no feedback condition.

Fourth, a mediational hypothesis was proposed: since a) feedback is described as one critical factor leading to accurate calibration and b) calibration accuracy has shown to beneficially influence performance, it is assumed that accurate calibration is an underlying variable mediating beneficial competential feedback effects on performance.

2 METHOD

The experimental condition with Competential feedback was compared to a Social feedback condition and to a Control condition without any feedback. Hypotheses were tested with a pre-post experimental design.

2.1 Sample and procedure

Participants were $N = 120$ secondary school students in grade 8 ($M = 13.99$ years; $SD = 0.33$; 65 females). The sample was drawn from four schools in Hamburg, Germany. In each class, students were randomly divided into three groups and allocated to either the Competential, the Social, or the No-feedback condition.

2.2 Feedback conditions

The three feedback conditions were Competential feedback, Social feedback, and No feedback (control). All feedback conditions were realized via feedback sheets. To make sure that intervention conditions did not differ in the amount of information given, all feedback sheets contained the same general information about scientific reasoning. This information was based on a competence model regarding scientific reasoning (Mayer, Grube, & Möller, 2008). Participants were told: "When planning experiments, you have to consider the

following steps: (1) At first, you say what you want to investigate. You name a variable. A variable is the factor which is going to be investigated and modified, i.e. varied in the experiment. (2) You say how you are going to measure the modifications of the variable. (3) You control other influences which you don't want to investigate, i.e. you keep all other variables constant. (4) You say with how many objects of investigation, how often or how long you want to run the experiment. (5) You think of possible mistakes or limitations in your planning. Then you plan a new and better experiment." These five levels of scientific reasoning were presented in a chart.

Intervention conditions and respectively the feedback sheets only varied relative to the feedback information: students' achievements were either evaluated and fed back in relation to the competence model (Competential), in relation to their classmates' achievements (Social), or merely given the form without any marks or comments (Control).

2.2.1 Competential feedback

In the Competential feedback condition students' achievements were evaluated in relation to the above described five levels of scientific reasoning (see section 2.2) (Mayer et al., 2008). Each of the five experiments planned by the students was rated by assessing which level of scientific reasoning had been considered. When the students had indicated what they wanted to investigate and had named a variable, for example, they received a "tick" to indicate that this level (1) was successful. Then, when they had indicated how they were going to measure the modifications of the variable, they received a "tick" for level 2, and so on. They had to move through the sequence from level 1 to 5 and could ultimately end with between zero to five ticks. When they had not completed a particular level then there was no "tick" assigned (i.e., it was unmarked). Students were told that they could improve in the next task by attending to the requirements indicated by the subsequent unmarked levels. When a

student for example had named a variable (level 1), explained how to measure the modifications of the variable (level 2) and controlled for other influences (level 3), levels 1-3 were “ticked” in the chart on the feedback sheet. If so, levels 4 and 5 would have been unmarked and the following advice provided: “(4) you need to say how many objects of investigation, how often or how long you want to run the experiment, and (5) you need to think of possible mistakes or limitations in your planning. Then you plan a new and better experiment.” The aim was to provide students with the level of competence at which they were performing and then to give them feedback aiming to advance them to the next levels of competence.

2.2.2 Social feedback

The Social feedback indicated the students’ achievement relative to their classmates’ achievements. Although the Social feedback sheet contained the five levels of scientific reasoning as general information (see section 2.2), there were no marks as to which levels the students had completed correctly or not. Instead, an overall grade of 1 to 5 was presented in a separate chart, where grade 1 equated to “very good” (“sehr gut”) and grade 5 equated to “unsatisfactory or inadequate” (“mangelhaft”). Students were told: “This chart shows how well you performed compared to your classmates” as the aim was to induce a sense of social feedback. This overall performance rating was based on which level of scientific reasoning students had reached: Grade 1 (“very good”) matched competence level 5, grade 2 matched competence level 4, and so on.

2.2.3 Control condition

In the No-feedback condition students were given the general information concerning the five levels of scientific reasoning (see section 2.2), but their performance was not

evaluated in any way. They were neither told their competence level nor their achievement in relation to their classmates' achievements.

2.3 Measures

2.3.1 Performance

Effects on students' performance in scientific reasoning were investigated by means of five paper-pencil-tasks developed by Mayer et al. (2008), asking students to individually plan scientific experiments. The tasks required the planning of experiments related to the students' everyday life situations and could be planned by the students without any task-specific prior knowledge (e.g., "Tom wants to give some roses to his mother as a present for her birthday. He has read in a gardening journal that roses would grow better in flower soil than in garden mould. Before giving the roses to his mom, he wants to investigate if this statement in the gardening journal is correct. Please plan an experiment by which the statement can be investigated scientifically. Describe the experiment as accurately as possible"). Answers were coded on a scale from competence levels 1-5 (as noted above) according to a coding rubric provided by Mayer et al. (2008).

2.3.2 Motivation

Motivation was measured using two subscales (3 items each using a 1-4 Likert scale), based on a questionnaire called "Short Scale of Intrinsic Motivation" or in German "Kurzsкала intrinsischer Motivation" (Wilde, Bätz, Kovaleva, & Urhahne, 2009). The two scales, Interest and Perceived Competence, were based on the self-determination theories by Deci and Ryan (2000). Three items of each of the subscales were adapted to the domain of scientific reasoning: Interest (e.g., "Planning the experiments was very interesting"), Perceived Competence (e.g., "I am satisfied with my performance at planning the

experiments“). Students’ answers could range from 1 = *disagree completely* to 4 = *agree completely*. Motivation was measured right before task 1 and after task 5. The pre-post measurement points for motivation were chosen to enable students to work on the tasks without cease. This aimed at inducing a natural learning environment.

2.3.3 Calibration

Two different measures of calibration were used, based on the research by Schraw et al. (1993). The first measure aimed to assess the direction of self-evaluative judgment errors. This calibration bias comprises the difference between true and estimated performance indicating over- or underconfidence. According to Schraw et al. (1993), the mean bias score was computed by subtracting true performance (i.e., the students’ levels of performance or mark relative to their classmates’, as indicated above) from estimated performance averaged for all measuring points. Students estimated their performance before tasks two, three and four answering the question how many correct answers they expected themselves to give in the next experiment on a scale from 0 to 5. Mean bias score therefore ranged from -5 to 5, with scores below zero indicating underconfidence, 0 indicating perfect calibration, and scores above zero indicating overconfidence.

The second calibration measure aimed to assess the accuracy of self-evaluative performance judgments. Calibration accuracy was calculated by the absolute value of the difference between estimated and true judgments averaged across measuring points. The accuracy score ranged from 0 to 5, with 0 indicating perfect accuracy and 5 indicating a lack of accuracy.

2.3.4 Pre-group differences

Although the students were randomly assigned to the three treatment groups, four measures were used to check whether there were initial differences between the groups. First, a measure of the underlying competence, a nonverbal reasoning scale of the KFT 4-12+R, was used (Heller & Perleth, 2000). The KFT 4-12+R is a German adaptation of the Cognitive Abilities Test developed by Thorndike and Hagen (1971). The *figure analogy* scale contains 25 items selected for 8th grade and these items were presented in a multiple choice, paper-and-pencil format. Second, two items were developed to measure Interest in Biology (as the task related to daily life biology examples; e.g., “I am interested in biology.”). Third, two items were developed to measure Experience in Planning Experiments (e.g., “I have already planned experiments like that before”). Fourth, performance on the first task (planning of experiments) was checked to ensure no pre-group differences prior to receiving feedback or not.

A manipulation check was also developed to evaluate the impact of the two feedback conditions. Six items from the “Teacher Reference Norm – Student View” or in German “Schülerperzipierte Lehrerbezugsnorm” (Schwarzer, Lange, & Jerusalem, 1982) were used. Three of these items related to the social reference norm (e.g., “The feedback sheets between the tasks showed me how I performed in relation to my classmates”), and three items related to the competential reference norm (e.g., “The feedback sheet between the tasks showed me what I had to consider in the following task”). These items were rated on a four-point rating scale (1 = *disagree completely* to 4 = *agree completely*). This feedback manipulation items were administered after all 5 tasks were completed and all feedback had been provided.

2.4 Procedure

All test instructions were read to the students by the experimenter to achieve comparability between the experimental conditions. First, the general reasoning questionnaire was conducted. Then, before starting to work on the tasks, all participants received a copy of the five scientific reasoning steps and each step was explained to them. This was followed by the motivational scales (Interest and Perceived Competence). At this point, the students planned the five experiments for which they had a maximum of 10 minutes each. After each experiment the performance was rated by means of a coding manual (Mayer et al., 2008) and the respective feedback sheet was handed out. Finally, control measures of Interest in Biology and Experience in Planning Experiments and the demographic items (sex, age, school) concluded the study. The study took 90 minutes. Ethics approval was granted by the Department of Education in Hamburg.

2.5 Methods of analysis

Across all 120 students, there were 3 students for Perceived Competence (2.5%), 4 for Bias (3.3 %), and 4 for Accuracy (3.3 %) who did not complete all items on these scales. Because of this small number of missing values, EM imputation by means of a missing value analyses (SPSS) was regarded as adequate. Coefficient alpha was used to estimate reliability, and *MANOVA* to evaluate differences in the dependent variables moderated by group membership. To ensure comparability and remove any difficulty-differences of the tasks, *z*-standardized means were used in subsequent performance analyses.

3 RESULTS

3.1 Estimates of reliability

The estimates of reliability were calculated for each scale: Performance tasks $\alpha = .90$, Motivation-Interest $\alpha = .87$, Motivation-Perceived Competence $\alpha = .83$; Calibration-Bias $\alpha = .74$; Calibration-Accuracy $\alpha = .74$; Control-Reasoning $\alpha = .85$; Control-Interest in Biology $\alpha = .66$; Control-Experience in Planning Experiments $\alpha = .76$; Manipulation-Feedback Competential $\alpha = .69$; and Manipulation-Feedback Social $\alpha = .85$. In all cases the estimates were sufficiently high to provide confidence to use total scores in subsequent analyses.

Inter-rater reliability was used to evaluate the consistency of the coding of the five tasks to the levels or marks. During the experiment coding was undertaken by the researcher as immediate feedback needed to be provided. After the experiment, two researchers independently coded the responses to each task according the rubric in the manual, and these two researchers were blind as to which condition the responses emanated from. The ratings between the two blind and one researcher were quite high across the five tasks ($\kappa = 0.66$ to $\kappa = 0.84$; Cohen, 1960). The chance-corrected inter-rater reliability of task 1-5 by the two blind raters ranged from $\kappa = 0.74$ to $\kappa = 0.96$ and therefore was sufficiently high to provide confidence in the consistency of the ratings (Landis & Koch, 1977). Where there were discrepancies, the two blind raters were requested to agree on a rating. These ratings entered main data analyses.

3.2 Manipulation check

To test whether the experimentally varied feedback was perceived as Competential and respectively as Social feedback, intervention conditions were compared concerning the competential reference norm scale and the social reference norm scale. A *MANOVA* showed a

statistically significant difference between intervention conditions in both scales (*Wilks Lambda* = .79; *Mult.F*(4,228) = 7.14; $p < .001$; $\eta^2 = .11$). Both Competential and Social feedback showed differences between the three experimental groups (Competential $F(2,115) = 4.36$; $p < .001$; $\eta^2 = .07$; Social $F(2,115) = 12.40$; $p < .001$; $\eta^2 = .18$). Post-hoc tests revealed that participants in the Competential feedback condition ($M_{CF} = 3.50$; $SD = 0.60$) perceived their feedback significantly more competential than participants of the Social ($M_{SF} = 3.03$; $SD = 0.81$) and the No-feedback condition ($M_{CC} = 3.24$; $SD = 0.72$), whereas participants in the Social feedback condition perceived their feedback significantly more social ($M_{SF} = 2.95$; $SD = 0.66$) than participants of the Competential ($M_{CF} = 2.13$; $SD = 0.84$) and the No-feedback condition ($M_{CC} = 2.78$; $SD = 0.88$). This provides evidence of the effect of the treatment, in this case the Competential, the Social, or No-feedback.

3.3 Control variables

A *MANOVA* did not show any statistically significant differences across the three feedback groups in the mean scores for Experience in Planning Experiments, Interest in Biology, and General Reasoning (*Wilks Lambda* = .93; *Mult.F*(6, 212) = 1.29; $p = .26$), nor were there any univariate differences for each of the three variables (Interest in Biology $F(2,118) = 2.21$; $p = .11$; Planning Experiments $F(2,118) = .31$; $p = .73$; Reasoning $F(2,118) = 1.77$; $p = .18$).

3.4 Main analyses

The main analyses concerned the impact of feedback variation on performance and motivation, (i.e., Interest and Perceived Competence), and Calibration. Means and standard deviations of the following analyses are all shown in Table 1.

Table 1

Means (standard deviations) concerning the dependent variables Performance (z-standardized), Interest, Perceived Competence, and Calibration, i.e. Bias and Accuracy of self-evaluative performance judgments

	Feedback		Control group
	competential (<i>n</i> = 40)	social (<i>n</i> = 40)	no-feedback (<i>n</i> = 40)
Performance			
Task 1 (Pretest)	0.25 (0.91)	-0.02 (0.98)	-0.23 (1.07)
Task 2	0.63 (1.00)	-0.30 (0.88)	-0.33 (0.82)
Task 3	0.73 (0.98)	-0.22 (0.79)	-0.52 (0.77)
Task 4	0.80 (0.94)	-0.30 (0.81)	-0.50 (0.70)
Task 5	0.72 (0.97)	-0.21 (0.84)	-0.50 (0.76)
Interest			
Pretest	3.15 (0.58)	3.15 (0.50)	3.00 (0.61)
Posttest	3.07 (0.73)	3.09 (0.76)	2.76 (0.89)
Perceived Competence			
Pretest	2.87 (0.55)	3.10 (0.44)	3.04 (0.53)
Posttest	2.83 (0.68)	2.83 (0.71)	2.44 (0.74)
Calibration			
Bias	0.17 (1.15)	1.30 (1.11)	1.45 (1.00)
Accuracy	1.13 (0.55)	1.47 (0.96)	1.59 (0.82)

Note: Performance was coded on competence levels 1-5; a higher number indicating a higher competence level. Interest and Perceived Competence items were rated on a four-point rating scale from 1 = *disagree completely* to 4 = *agree completely*. Calibration bias is the difference between true and estimated performance indicating over- or underconfidence; mean bias score ranged from -5 to 5, with 0 indicating perfect calibration. Calibration accuracy is the absolute value of the difference between true and estimated performance; the accuracy score ranged from 0 to 5, with 0 indicating perfect accuracy.

3.4.1 Performance

Participants' performance in scientific reasoning was examined by means of five tasks (Mayer et al., 2008). The first measuring point was set before the feedback information was distributed; it thus can be considered baseline information or pretest performance. Measuring points 2-5 comprised four posttest measuring points. A repeated measures *MANOVA* with the factors *time* (t1-t5) and *feedback* was used to analyze feedback effects on posttest performance.

There was a statistically significant time by feedback interaction (*Wilks Lambda* = .81; *Mult.F*(8,228) = 3.23; $p = .002$; $\eta^2 = .10$).

There was not a significant difference between the three groups at T1 (pretest or baseline) (Roy-Bargmann $F(2,117) = 2.42$; $p = .09$; $\eta^2 = .04$), but there were appreciable differences at the subsequent three time points (Figure 1, T2 $F(2,116) = 13.06$; $p < .001$; $\eta^2 = .20$; T3 $F(2,115) = 8.86$; $p < .001$; $\eta^2 = .29$; T4 $F(2,114) = 7.04$; $p < .001$; $\eta^2 = .33$; T5 $F(2,113) = 1.57$; $p = .21$; $\eta^2 = .27$). Students in the Competential feedback condition increased their performance between T1 to T4, whereas students in the Social and No feedback condition changed hardly at all. The effect-size change between T1 and T5 is .49 for Competential, -.21 for Social, and -.30 for Control. *Z*-standardized means (*t*-transformed) of the performance tasks (1-5) are shown in Figure 1.

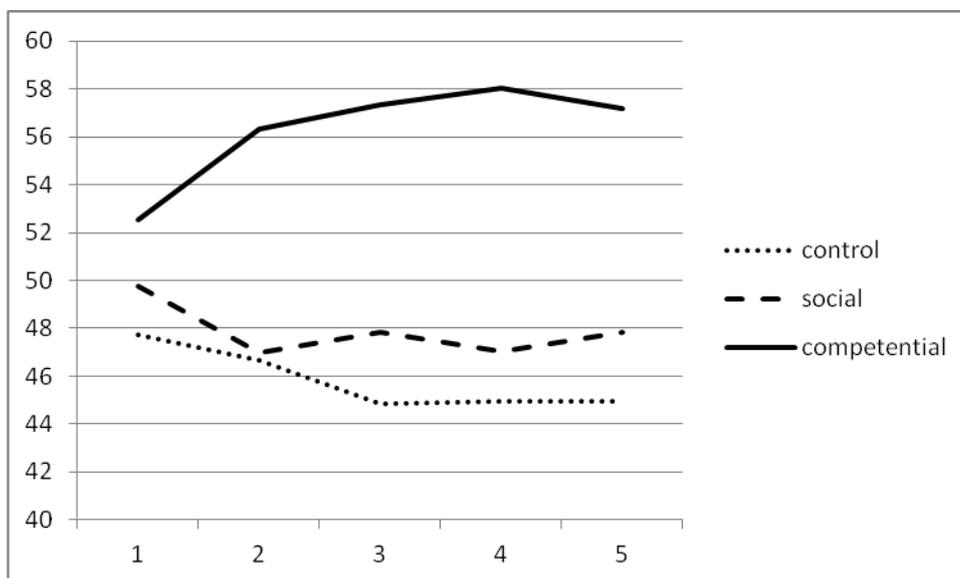


Figure 1. *T*-transformed means of the z -standardized Performance tasks (1-5).

3.4.2 Motivation

A repeated measures *MANOVA* was also used to detect differences in motivation, Interest and Perceived Competence. There were overall differences over time (*Wilks Lambda* = .85; *Mult.F*(2,116) = 0.86; $p < .001$), but Interest did not change over time ($F(2,117) = 0.74$; $p = .45$) whereas Perceived Competence changed over time ($F(2,117) = 3.16$; $p < .001$; $\eta^2 = .09$). The means show that for the Competential feedback condition ($d = -.07$) Perceived Competence was stable over time, whereas it decreased for the Social ($d = -.46$) and even more so for the Control condition ($d = -.94$) (Table 1).

3.4.3 Calibration

Concerning the mean bias (i.e., direction of self-evaluative judgment errors), a *MANOVA* showed that there were differences across the three groups (*Wilks Lambda* = .74; *Mult.F*(4,232) = 9.41; $p < .001$; $\eta^2 = .14$), and both Bias ($F(2,117) = 16.81$; $p < .001$; $\eta^2 = .22$) and Accuracy ($F(2,117) = 3.96$; $p = .02$; $\eta^2 = .06$) contributed to this overall difference.

Although participants of all intervention groups were overconfident, planned contrasts revealed that there was stronger bias in the Social feedback ($t(2,117) = 4.67$; $p < .001$; $d = .86$) and the No-feedback condition ($t(2,117) = 5.31$; $p < .001$; $d = .98$) compared to the Competential feedback condition. Mean bias did not differ significantly between the Social feedback vs. the No-feedback condition ($t(2,117) = 0.64$; $p = .52$; $d = .12$).

With regard to calibration accuracy, self-evaluative performance judgments in the Competential feedback condition were significantly more accurate than in the Social feedback condition ($t(2,117) = 1.88$; $p = .03$; $d = .35$) and in the No-feedback condition ($t(2,117) = 2.55$; $p = .01$; $d = .47$). Calibration accuracy did not differ significantly between the Social feedback vs. the No-feedback condition ($t(2,117) = 0.67$; $p = .50$, $d = .12$).

3.5 Mediation analyses

The next analysis investigated a potential underlying variable of the positive effect of Competential feedback on performance. On the one hand, it included linear regression analyses with regard to the impact of calibration accuracy on performance. On the other hand, it was examined whether calibration accuracy mediated the positive impact of Competential feedback on performance. Therefore, regression analyses were performed. Performance was the mean of *z*-standardized participants' posttest performance (t2-5). The calibration accuracy coefficient as the mean of the absolute value difference between estimated and true performance judgments entered the analyses. To be consistent in the direction of the effects, this coefficient was recoded before entering the analyses so that a higher number indicated higher accuracy. The Competential feedback coefficient comprised the contrast between the Competential (2) and the Social (-1) and No-feedback (-1) conditions.

First, the independent factor "feedback" had a significant effect on the dependent variable "performance", $\beta = .57$, $t(118) = 7.59$, $p < .001$. The experimentally varied feedback information explained a significant proportion of variance of performance, $R^2 = .33$, $F(1,118) = 57.64$, $p < .001$. Second, according to the regression analyses, the independent factor "feedback" was a significant predictor of the mediator variable "calibration accuracy", $\beta = .23$, $t(118) = 2.52$, $p = .01$. The experimentally varied feedback information explained a significant proportion of variance of calibration accuracy, $R^2 = .05$, $F(1,118) = 6.36$, $p = .01$. Third, concerning the impact of calibration accuracy on performance, the accuracy of self-evaluative performance judgments predicted the dependent variable "performance", $\beta = .54$, $t(118) = 6.99$, $p < .001$. Calibration accuracy explained a significant proportion of variance of performance, $R^2 = .29$, $F(1,118) = 48.88$, $p < .001$.

Therefore, calibration accuracy can be a mediator for the impact of Competential feedback on performance as the present study showed. This is due to the fact that in all three

regression analyses, the predictor significantly predicts the criterion, which is needed to be met to prove mediation (Baron & Kenny, 1986).

The Sobel test (Sobel, 1982) was used to investigate whether the mediator variable calibration accuracy significantly conveys the influence of feedback on performance. It is a significance test of indirect impact (see MacKinnon, Warsi, & Dwyer, 1995). The Sobel test reached statistical significance indicating that the mediation was statistically significant (*SEb Calibration accuracy Feedback.Performance Calibration accuracy = .07; b Calibration accuracy Feedback.Performance Calibration accuracy = .33; tSobel = 2.19; p = .03*).

4 DISCUSSION

4.1 Conclusions

This study aimed to investigate the effectiveness of providing competence based feedback to students and see the effects relative to social and no feedback. The competence model was used as an external standard to which students' achievements were compared and fed back to them. Thus, the competential feedback is a theoretically derived extension of a criterion-related feedback. Competential feedback was assumed to inform students about their correct answers (how am I going?), to be task-specific, to elucidate goals and to include cues on the current state of their learning process (where am I going?) as well as on how to improve their task performance (where to next?). These feedback aspects have been identified as critical in making feedback powerful.

In line with the first hypothesis, competential feedback led to a significantly improved task performance when compared with the social and the no-feedback condition. Indeed social feedback hardly differed from no-feedback at all. Competential feedback also induced an increase in effect sizes over time whereas social and no-feedback led to a sudden performance decrease after the first feedback was given. Although the performance decrease

effect for social feedback slightly recovered and then stabilized over time, students' performance in the social feedback condition remained significantly lower than in the competential feedback condition throughout all tasks.

Contrary to our second hypotheses, interest was not influenced by the feedback conditions. But there was a feedback effect on perceived competence: Competential feedback was able to maintain students' perceived competence, and there was a decrease of perceived competence when students were compared to their classmates or to those who did not receive any feedback at all. Thus, there was a beneficial effect when comparing competential feedback to the other two intervention conditions. One could assume that the decrease of perceived competence in the social and the no-feedback condition accounted for the performance decrease in these groups. However, this hypothesis needs to be investigated in future studies.

In considering the third hypothesis regarding calibration bias and accuracy, expected effects were confirmed: performance judgments of the students in the competential feedback condition were a) less biased and b) more accurate than in the social feedback or in the no-feedback condition. This finding is striking, especially considering the short intervention period. Hence, this study was able to give empirical evidence to the theoretically described impact of effective external feedback leading to effective internal feedback (Butler & Winne, 1995; Stone, 2000).

Fourth, there was a significant indirect effect of calibration: calibration accuracy was found to be a partial mediator of the feedback impact on performance. Competential feedback led to more accurate performance judgments which in turn contributed to the beneficial performance effects. This finding emphasizes a) the importance of students' accurate performance judgments and b) the possibility of effective feedback methods in achieving such accurate performance judgments.

In summary, this study shows a beneficial impact on performance, perceived competence and calibration. Beyond that, calibration was identified as one underlying variable mediating the positive impact of competential feedback on performance.

4.2 Limitations and future studies

These results have to be interpreted with regard to the following limitations: In this design, there was only a short intervention period. Thus, some social feedback effects might not have been revealed. However, the identified feedback effects, in spite of the short intervention period, still led to marked feedback effects.

Furthermore, prompting effects (Hoffman & Spataritu, 2008; Hübner, Nückles, & Renkl, 2010) and feedback effects are confounded since feedback was given immediately before the performance tasks. Further studies should investigate feedback effects by giving more time between feedback and performance tasks.

With regard to the motivational variables, findings can only be generalized to a limited extent: first, interest and perceived competence should not only be measured once before and after all performance tasks, but also separately before and after each task. Thus, feedback impact on motivational variables can be investigated over time. Second, motivational variables should be measured using a longer version of motivational scales to capture more feedback effects.

Further studies could focus on investigating a combination of social and competential feedback that would be applicable to classroom assessment. To clearly explain the feedback effects on performance other mediators or moderators need identification. Possible moderator effects could be the students' pre-conditions, such as high-achieving vs. low-achieving students, or under- vs. overconfident students (Labuhn et al., 2010). Another promising approach would be to link the competential feedback with the research done on worked-out

examples (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 1997). Worked-out examples comprise a problem, solution steps, and the final solution itself; this learning mode has shown to be effective in the learning process, although showing differential effects of learners' pre-knowledge (Salden, Alevén, Schwonke, & Renkl, 2010).

4.3 Implications

The innovative feedback condition “competential feedback” was theoretically derived and based on former research findings concerning feedback effects. It was compared to commonly used feedback in schools (mark 1-5). Thus, implications can be drawn for educational research and practice.

The theoretical contribution of competential feedback comprises the combination of two research lines such as feedback findings and research on competence models. This study revealed direct beneficial competential feedback effects on performance, perceived interest, and accuracy of performance judgments, but also looked at indirect, underlying effects. This leads to educational implications: having empirically identified a) beneficial effects of competential feedback on learning outcomes and b) accuracy of performance judgments as one underlying variable of feedback effectiveness on learning outcomes, could have great benefit for schools.

5 REFERENCES

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271. doi:10.1037/0022-0663.84.3.261
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181-214. doi:10.3102/00346543070002181
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238. doi:10.3102/00346543061002213
- Baron, K. E., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi:10.1037/0022-3514.51.6.1173
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74. doi:10.1080/0969595980050102
- Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*, 269-290. doi:10.3200/JEXE.73.4.269-290
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research, 65*, 245-281. doi:10.3102/00346543065003245
- Chen, P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 79-92. doi:10.1016/j.lindif.2003.08.003

- Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A schoolbased program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools, 41*, 537-550. doi:10.1002/pits.10177
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. doi:10.1177/001316446002000104
- Connell, M. W., Sheridan, K., & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126-155). Cambridge, MA: Cambridge University Press.
- Covington, M.V. (2001). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*, 171-200. doi:10.1146/annurev.psych.51.1.171
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227-268. doi:10.1207/S15327965PLI1104_01
- Duijhouwer, H., Prins, F. J., & Stokking, K. M. (2011). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction*. Advance online publication. doi:10.1016/j.learninstruc.2011.10.003
- Harks, B., Rakoczy, K., Hattie, J., Klieme E., & Besser, M. (2011). *Indirect feedback effects on mathematics achievement*. Manuscript submitted for publication.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik [Competence and competence diagnosis]. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* [Performance and assessment of performance] (pp. 127-143). Berlin: Springer-Verlag.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi:10.3102/003465430298487
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeits-Test für 4.-12. Klassen, Revision (KFT 4-12+R)* [Cognitive Ability Test for grades 4 to 12, revised version]. Göttingen: Beltz.
- Hoffman, B., & Spatariu, A. (2008). The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology*, 33, 875-893. doi:10.1016/j.cedpsych.2007.07.002
- Horgan, D. (1990). Competition, calibration, and motivation. *Teaching Thinking and Problem Solving*, 12, 5-10.
- Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, 20, 18-29. doi:10.1016/j.learninstruc.2008.12.001
- Kitsantas, A., & Zimmerman, B. J. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning*, 1, 202-212. doi:10.1007/s11409-006-9000-7
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht [Performance assessment and competence modeling in mathematics instruction]. [Special Issue on "Kompetenzmodellierung"] *Zeitschrift für Pädagogik*, 56, 64-74.

- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3-22). Göttingen: Hogrefe.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen [Competence models for assessing individual learning outcomes and evaluating educational outcomes]. *Zeitschrift für Pädagogik*, 52, 876-903. Retrieved from http://www.pedocs.de/frontdoor.php?source_opus=4493
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284. doi:10.1037//0033-2909.119.2.254
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008): Current issues in competence modeling and assessment. *Journal of Psychology*, 216, 61-73. doi:10.1027/0044-3409.216.2.61
- Krampen, G. (1987). Differential effects of teacher comments. *Journal of Educational Psychology*, 79, 137-146. doi:10.1037/0022-0663.79.2.137
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134. doi:10.1037/0022-3514.77.6.1121
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232. doi:10.3102/00346543047002211

- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173-194. doi:10.1007/s11409-010-9056-2
- Landis, G., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62. doi:10.1207/s15327906mbr3001_3
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung [Competence model of Scientific Inquiry]. In U. Harms, & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* [Research in Teaching and Instruction in Biology Education] (pp. 63-79). Innsbruck: Studienverlag.
- Mischo, C., & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung [Educational goals of teachers and individual frames of reference for performance assessment]. *Zeitschrift für Pädagogische Psychologie*, 9, 139-151.
- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

- Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*, 310-322. doi:10.1016/j.learninstruc.2006.07.003
- Nolen, S. B. (1996). Why study? How reasons for learning influence strategy selection. *Educational Psychology Review, 8*, 335-355. doi:10.1007/BF01463938
- Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving: implications of using different forms of assessment. *The Journal of Experimental Education, 65*, 213-228. doi:10.1080/00220973.1997.9943455
- Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction. Grading and feedback in mathematics classrooms. *Journal of Psychology, 216*, 110-123. doi:10.1027/0044-3409.216.2.111
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1-29. Retrieved from <http://csjarchive.cogsci.rpi.edu/1997v21/index.html>
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung [Frame of references and the assessment of learning outcomes]. In W. Schneider, & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* [Handbook of educational psychology] (pp. 178-186). Göttingen: Hogrefe.
- Rheinberg, F., & Krug, S. (1999). *Motivationsförderung im Schulalltag* [Enhancing motivation in school]. Göttingen: Hogrefe.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144. doi:10.1007/BF00117714
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5*, 77-85. doi:10.1080/0969595980050104

- Salden, R., Alevan, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, 38, 289-307. doi: 10.1007/s11251-009-9107-8
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, 18, 455-463. doi:10.1006/ceps.1993.1034
- Schwarzer, R., Lange, B., & Jerusalem, M. (1982). Die Bezugsnorm des Lehrers aus der Sicht des Schülers [Teachers' reference norms in the perception of pupils]. In F. Rheinberg (Ed.), *Bezugsnormen zur Schulleistungsbewertung* (pp. 161-172). Düsseldorf: Schwann.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189. doi:10.3102/0034654307313795
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290-312). Washington: American Sociological Association.
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12, 437-475. doi:1040-726X/00/1200-0437
- Thorndike, R. L., & Hagen, E. P. (1971). *Cognitive Abilities Test*. Boston: Houghton Mifflin.
- Van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58, 263-272. doi:10.1016/j.compedu.2011.07.020
- Vollmeyer R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15, 589-602. doi:10.1016/j.learninstruc.2005.08.001

- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe & Huber.
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM) [Testing a short scale of intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 31-45. Retrieved from http://www.ipn.uni-kiel.de/zfdn/pdf/15_Wilde.pdf
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551-572. doi:10.1016/S0361-476X(02)00006-1
- Wollenschläger, M., Möller, J., & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken [Effects of competential feedback on scientific reasoning]. *Zeitschrift für Pädagogische Psychologie*, 25, 197-202. doi:10.1024/1010-0652/a000040

ZUSAMMENFASSUNG

MEDIERT DIE GENAUIGKEIT DER SCHÜLERSELBSTEINSCHÄTZUNG KOMPETENZIELLE FEEDBACKEFFEKTE AUF PERFORMANZ?

Feedback ist definiert als die Information, die einer Person während oder nach einer Aufgabenbearbeitung dargeboten wird, um regulierend auf die zukünftige Aufgabenbearbeitung einzuwirken (Narciss, 2008). Feedback im Unterricht ist mit einer Effektgröße von $d = 0.79$ eine essenzielle Einflussgröße effektiver Lehr-/Lernprozesse (basierend auf 12 Meta-Analysen, 196 Studien und 6972 Effektgrößen; Hattie, 2009). Allerdings zeigt sich in diesen Meta-Analysen eine beträchtliche Variation der Effektgrößen und nicht alle Feedbacktypen erweisen sich als lernförderlich (Bangert-Drowns et al., 1991; Hattie, 2009; Kluger & DeNisi, 1996). Somit ist weitere Forschung nötig, die speziell die Effektivität von Feedback im Unterricht und im Lernprozess untersucht (Hattie, 2009). Shute (2008) schlägt diesbezüglich die Untersuchung aufgabenspezifischen Feedbacks sowie die Untersuchung von Mediationseffekten vor; mithilfe der Mediationseffekte sollen zugrundeliegende Variablen identifiziert werden, die zur Effektivität von Feedback beitragen.

Die Kompetenzforschung bietet die Möglichkeit, Schülerleistungen in Relation zu Stufen eines domänenspezifischen Kompetenzmodells zu bewerten (Koeppen et al., 2008). Die Rückmeldung der erreichten und (noch) zu erreichenden Kompetenzstufen ist definiert als kompetenzielles Feedback. Dieses Feedback erweist sich als lern- und motivationsförderlich, wenn es mit einer Kontrollbedingung ohne Feedback verglichen wird (Wollenschläger et al., 2011). Bisher wurde jedoch noch nicht untersucht, inwiefern kompetenzielles Feedback anderen Feedbackinformation über- oder unterlegen ist.

Die vorliegende Studie greift die oben genannten Forschungsaufgaben auf: erstens wird mit kompetenziellem Feedback ein aufgabenspezifisches Feedback untersucht, zweitens werden dessen Effekte auf Performanz, Motivation und Genauigkeit der

Schülerselbsteinschätzung vergleichend zu im Schulkontext gängiger Feedbackinformation untersucht; drittens wird die Genauigkeit der Schülerselbsteinschätzung als potenzielle zugrundliegende Variable und Mediator positiver Feedbackeffekte auf Performanz untersucht.

Kompetenzielles Feedback ist als Leistungsbewertung in Relation zu Stufen eines domänenspezifischen Kompetenzmodells definiert. Effekte dieses kriterialen Feedbacks auf Performanz, Motivation (Kompetenzerleben) und Genauigkeit der Schülerselbsteinschätzung wurden untersucht. In einem Pre-Posttest-Design wurden Schülerinnen und Schüler der Sekundarstufe 1 ($N = 120$) entweder einer Bedingung mit kompetenziellem Feedback oder einer Bedingung mit sozialem Feedback (Einstufung der Schülerleistungen in Relation zum Klassendurchschnitt) oder einer Vergleichsbedingung ohne Feedback randomisiert zugeteilt.

Lernende in der kompetenziellen Feedbackbedingung schnitten signifikant besser ab, nahmen sich selbst kompetenter wahr und schätzten ihre eigene Leistung signifikant genauer ein als Lernende mit sozialem oder keinem Feedback. Die Genauigkeit der Performanzeinschätzung konnte als partieller Mediator des Feedbackeffekts auf Performanz identifiziert werden: kompetenzielles Feedback führte zu genaueren Selbsteinschätzungen der Lernenden, diese wiederum trugen zu einer Performanzsteigerung bei.

Folglich gibt diese Studie Hinweise darauf, dass erstens theoretisch fundiertes Feedback basierend auf einem Kompetenzmodell im Lernprozess effektiv ist und es zweitens sozialem Feedback (wie es häufig im Schulalltag eingesetzt wird) überlegen ist. Darüber hinaus zeigt sich durch die Untersuchung des Mediationseffekts, dass die Genauigkeit der Schülerselbsteinschätzung eine zentrale Rolle für die Effektivität von Feedbackinformationen spielt.

STUDIE 3

EXTERNAL VERSUS INTERNAL COMPETENTIAL FEEDBACK: DIFFERENTIAL EFFECTS ON PERFORMANCE AND CALIBRATION?

Wollenschläger, M., Möller, J., & Harms, U. (2012). *External versus Internal
Competential Feedback: Differential Effects on Performance and Calibration?*

Manuskript eingereicht zur Publikation.

ABSTRACT

Competential feedback is defined as the evaluation of students' achievements in relation to defined competency levels within a domain-specific competence model. This study aimed at investigating differential effects of this criterion-related feedback on performance and calibration depending on the source of the feedback information (external vs. internal). In a pre-post-follow-up experimental design, randomly assigned secondary school students ($N = 75$) either received external competential feedback (i.e., given by the teacher) or internal competential feedback (i.e., they self-assessed their performance relative to the competency levels) or no feedback information at all (control group). Students in the external feedback condition showed a significantly better task performance in planning scientific experiments and were also more accurate in their self-evaluative performance judgements than those students in the internal feedback and the control condition. This calibration accuracy partially mediated the differential feedback effects on performance: external feedback led to more accurate self-evaluations which in turn accounted for improved students' performance. This study underlines a) the benefit of external competential feedback and b) the importance of accurate self-evaluations in self-assessment and feedback reception.

1 INTRODUCTION

Butler and Winne (1995) differentiated between external feedback (i.e., provided by the teacher or an external source) versus internal feedback (i.e., generated by the students) through self-assessment. So far, competential feedback has been examined when externally provided by a teacher. External competential feedback evaluating students' performance relative to defined competency levels has shown to be effective when compared to either no feedback (Wollenschläger, Möller, & Harms, 2011) or to social feedback (the students were given marks relative to their classmates' achievements, see Wollenschläger, Hattie, Möller, & Harms, 2011): Competential teacher feedback improved students' performance and led to more accurate self-evaluative performance judgements; this calibration accuracy mediated the positive effect of competential feedback on performance (Wollenschläger et al., 2011a). Effects of students' self-assessments relative to defined competency levels (i.e., internal competential feedback) have not yet been investigated.

Self-assessment has shown to have a beneficial impact on performance (e.g., Stiggins & Chapuis, 2006) and also on the development of metacognitive skills (e.g., Cooper, 2006). Self-assessment based on defined rubrics has been studied (Andrade & Valtcheva, 2009), but a) only compared to a control condition without self-assessment-intervention and b) only with respect to performance improvement. Effects of using the same rubric for teachers' and students' and their effects on both, performance and metacognitive processes, need to be examined.

When using competency levels as defined rubrics, the question is if the above described beneficial effects of competential teachers' feedback can be found for self-assessment as well. Thus, this study aims at investigating differential competential feedback effects on students' performance and their self-evaluative performance judgements dependent on the feedback source (i.e., teacher vs. self-assessment).

1.1 Competential feedback

Competential feedback is defined as the evaluation of students' achievements in relation to defined competency levels within a domain-specific competence model (Wollenschläger et al., 2011b). Therefore, competential feedback combines research findings on feedback effectiveness and research on competence models.

Research on feedback effectiveness has its origin in the fact that providing feedback in classrooms is found to have a very powerful impact on students' learning (Hattie, 2009), but feedback effects are not necessarily beneficial and reported effect sizes can vary considerably (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie, 2009; Kulhavy, 1977). Therefore, meta-analyses have identified certain aspects making feedback effective on learning: It has been shown that feedback should focus on correct answers, needs to be task-specific, should elucidate specific goals telling learners the what, how and why of the task, and most importantly, should include cues on the current state of the learning process as well as on the how to improve task performance (Black & Wiliam, 1998; Hattie, 2009, 2012; Hattie & Gan, 2011; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mory, 2004; Sadler, 1989, 1998; Shute, 2008). As summarized in Hattie and Timperley's feedback model (2007, p. 87), effective feedback has to answer one of the three questions: "Where am I going, How am I going, and Where to next". These effective feedback attributes have been linked with research on competence models.

In the last decade, the term "competence" has been increasingly used in educational research and psychology (Connel, Sheridan, & Gardner, 2003; Klieme, Hartig & Rauch, 2008; Klieme, Bürgermeister, Harks, Blum, Leiß, & Rakoczy, 2010; Weinert, 2001). Competencies are defined as context-specific cognitive ability constructs that relate to situations and demands in domains concerning specific areas of learning (Klieme & Leutner, 2006). Competence models comprise the internal structure as well as different levels of

competencies with regard to domain-specific performance and take into account developmental processes (Koeppen, Hartig, Klieme, & Leutner, 2008).

Evaluating students' achievements in relation to defined competency levels within a domain-specific competence model combines these two research lines: Feedback which is based on competence models, illustrating reached and desired competency levels, focuses on correct answers, is task-specific, goal-centered and informative as to how to improve task performance since it includes information about the actual as well as the potential performance, elucidating a rubric of desired performance (Wollenschläger et al., 2011a). In summary, competential feedback seems to address all aspects of effective feedback (where should the student go, how does the student go, and where does the student go to next) (Hattie & Wollenschläger, 2011a).

Since feedback effects depend on students' motives and goals (Butler & Winne, 1995; Nolen, 1996), feedback studies need to take into account underlying variables mediating feedback influences on performance (e.g., Harks, Rakoczy, Hattie, Klieme & Besser, 2011; Shute, 2008). Accuracy of self-evaluated performance judgments has been suggested (Stone, 2000) and examined (Labuhn, Zimmerman, & Hasselhorn, 2010; Wollenschläger et al., 2011a) as a potential mediator of feedback influences. The magnitude of consent between learners' true and self-estimated test performance is defined as calibration accuracy (Schraw, Potenza, & Nebelsick-Gullet, 1993). Divergence in true and perceived performance displays either over- or underconfidence. Overconfidence is supposed to impede performance whereas accurate calibration leads to effective learning strategies and performance increase (e.g., Bol, Hacker, O'Shea, & Allen, 2005; Chen, 2002; Kruger & Dunning, 2009; Labuhn et al., 2010).

Thus, studies investigating the effectiveness of competential feedback also took into account mediational variables.

1.2 External competential feedback

External competential feedback is defined as feedback given by a teacher relative to a competence model. In previous studies, a teacher (i.e., the researcher, to keep the teacher variable constant) evaluated students' achievements relative to a domain-specific competence model; the domain was scientific reasoning in science classes (Wollenschläger et al., 2011b; Wollenschläger et al., 2011a). Effects on students' performance in scientific reasoning and calibration accuracy were investigated. Competential feedback which is given by a teacher based on the reached and not yet reached levels of a competence model has shown to improve students' learning outcomes, both performance and calibration accuracy, compared to a control condition without any feedback (Wollenschläger et al., 2011b) and compared to social feedback (the students were given marks relative to their classmates' achievements) (Wollenschläger et al., 2011a). Beyond that, calibration accuracy as an underlying variable mediating the feedback effects on performance has been identified: the more accurate students self-evaluated their performance, the better they performed (Wollenschläger et al., 2011a).

So far, external competential feedback effects on students' learning outcome have been investigated when feedback information was provided formatively (i.e., during the learning process and between each of the performance tasks) (Wollenschläger et al., 2011a,b). When providing external feedback between each of the tasks, prompting (Hoffman & Spataru, 2008; Hübner, Nückles, & Renkl, 2010) and feedback effects cannot be differentiated. Prompting is the information which is provided immediately before task completion, aiming at improving learning outcome (e.g., Hoffman & Spataru, 2008). To control for the prompting effect in explaining performance increase, external feedback effects need to be studied when feedback information is not provided immediately before each task.

Thus, our study aims at extending existing research on external competential feedback a) by including a follow-up-measure of performance without previous feedback information and b) by combining it with research on internal feedback (i.e., students' self-assessments).

1.3 Internal feedback – Students' self-assessments

In formative assessment literature (e.g., Bennett, 2011; Sadler, 1998), self-assessment feedback is seen as a crucial factor leading to students' performance increase: "Self-assessment by pupils [...] is in fact an essential component of formative assessment. When anyone is trying to learn, feedback about the effort has three elements: recognition of the desired goal, evidence about the present position, and some understanding of a way to close the gap between the two. All three must be understood to some degree by anyone before he or she can take action to improve learning." (Black & Wiliam, 1998, p. 143).

There is a large body of research on the effects of students' self-assessments on learning outcomes, especially in higher education (Boud & Falchikov, 1989, 2006; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989; Lew, Alwis, & Schmidt, 2010; Nulty, 2011). But also in primary (e.g., Towler & Broadfoot, 1992) and secondary (e.g., White & Frederiksen, 1998) school, self-assessment has shown to improve students' performance, for example their writing (Andrade, Du, & Wang, 2008; Ross, Rolheiser, & Hogaboam-Gray, 1999) or their mathematic problem-solving abilities (Ross, Hogaboam-Gray, & Rolheiser, 2002). Some authors conjectured that "students can be their own source of feedback, given the appropriate conditions and supports" (Goodrich Andrade & Boulay, 2003, p. 22). Based on this hypothesis, Goodrich Andrade and colleagues examined defined rubrics as standards-referenced assessment tools: Having students self-assessing their performance relative to defined rubrics led to an increased writing performance when compared to a control group without the self-assessment intervention

(Andrade, H. & Valtcheva, 2009; Goodrich Andrade & Boulay, 2003). These studies did not compare self-assessment with teachers' feedback and only focused on students' performance. However, self-assessment has shown to influence other learning variables as well: Self-assessment has shown to induce students' self-reflections (Cyboran, 2006) and to positively influence the development of metacognitive skills (e.g., Cooper, 2006), thereby helping students to monitor their learning (Bangert-Drowns et al., 1991; Black & Wiliam, 1998; Butler & Winne, 1995; White & Frederiksen, 1998). Thus, it seems worthwhile to examine self-assessment effects also concerning calibration accuracy: Does internal competential feedback (i.e., self-assessment) lead to more accurate self-evaluations as it has been shown for competential teachers' feedback?

The present study extends existing research on defined self-assessment rubrics, a) by using competency levels as defined rubrics, b) by comparing it to teacher feedback based on the same rubrics, and c) by investigating an underlying variable (i.e., calibration accuracy) mediating the feedback effects on performance.

1.4 The present study

This study brings together two feedback types, competential teacher feedback and student self-assessment. Competential feedback effects on learning outcomes dependent on the feedback source (i.e., external or internal) are investigated. Competential teachers' feedback has been identified as an effective means of improving students' performance and calibration (Wollenschläger et al., 2011a).

First, the beneficial performance effect needs to be replicated in a setting where the feedback information is not provided immediately before each task completion. This is because otherwise prompting and feedback effects are confounded. Thus, a follow-up measuring point without previously providing feedback information is included.

Second, existing research on competential feedback as well as on self-assessment is extended by comparing external teachers' with internal students' feedback using the same assessment criteria (i.e., competency levels) compared to a control group without any feedback information. This design allows for examining potential differential effects on performance and calibration as a function of the feedback source. Compared to the control group without any feedback information, it is a) hypothesized that both, competential teachers' and students' feedback, improve task performance because the used assessment criteria (i.e., the reached and not yet reached competency levels) provide information on all feedback questions ("Where am I going, How am I going, and Where to next"). It is b) hypothesized that both, external as well as internal competential feedback, improve calibration accuracy compared to the control group without any feedback information because students are able to classify themselves according to the competency levels. To see one's own reached and desired competency levels should lead to less biased and more accurate self-evaluations in both feedback conditions.

Third, mediational effects are studied: Based on findings concerning external competential feedback effects, it is assumed that accurate self-evaluations lead to improved performance, also for internal competential feedback. Thus, it is hypothesized that increased calibration accuracy is a mediator of competential feedback effects on performance, independent of the feedback source.

2 METHOD

The experimental condition with External competential feedback given by the teacher (i.e., the researcher) was compared to an Internal competential feedback condition where students self-evaluated their performance relative to the competency levels, and to a Control

condition without any feedback. Hypotheses were tested with a pre-post-follow-up experimental design.

2.1 Sample and procedure

Participants were $N = 75$ secondary school students in grade 8 ($M = 14.55$ years; $SD = 0.66$; 35 females). The sample was drawn from four classes of two schools in Hamburg, Germany. In each class, students were randomly divided into three groups and allocated to either the External, the Internal, or the Control condition.

2.2 Feedback conditions

The three feedback conditions were External feedback, Internal feedback, and No feedback (control group). All feedback conditions were realized via feedback sheets. To make sure that intervention conditions did not differ in the amount of information given, all feedback sheets contained the same general information about scientific reasoning. This information was based on a competence model regarding scientific reasoning (Mayer, Grube & Möller, 2008). Participants were told: “When planning experiments, you have to consider the following steps: (1) At first, you say what you want to investigate. You name a variable. A variable is the factor which is going to be investigated and modified, i.e. varied in the experiment. (2) You say how you are going to measure the modifications of the variable. (3) You control other influences which you don’t want to investigate, i.e. you keep all other variables constant. (4) You say with how many objects of investigation, how often or how long you want to run the experiment. (5) You think of possible mistakes or limitations in your planning. Then you plan a new and better experiment.” These five levels of scientific reasoning were presented in a chart.

Intervention conditions and respectively the feedback sheets only varied relative to the feedback source: students' achievements relative to the described competency levels were either evaluated and fed back by the teacher or by the students themselves. In the Control condition students were merely given the form without any comments.

2.2.1 External competential feedback

In the External feedback condition students' achievements were evaluated in relation to the above described five levels of scientific reasoning (see section 2.2) (Mayer et al., 2008). Each of the three experiments planned by the students was rated by the teacher (i.e., the researcher, to keep the teacher variable constant) assessing which level of scientific reasoning had been considered. When the students had indicated what they wanted to investigate and named a variable, for example, they received a "tick" to indicate that this level (1) was successful. Then, when they had indicated how they were going to measure the modifications of the variable they received a "tick" for level 2, and so on. They had to move through the sequence from level 1 to 5 and could ultimately end up with ticks between zero to five. When they had not completed a particular level then there was no "tick" assigned (i.e., it was unmarked). Students were told that they could improve in the next task by attending to the requirements indicated by the subsequent unmarked levels. When a student, for example, had named a variable (level 1), explained how to measure the modifications of the variable (level 2) and controlled for other influences (level 3), level 1-3 were "ticked" in the chart on the feedback sheet. If so, levels 4 and 5 would have been unmarked and the following advice provided: "(4) you need to say how many objects of investigation, how often or how long you want to run the experiment", and "(5) you need to think of possible mistakes or limitations in your planning. Then you plan a new and better experiment." The aim was to provide students

with the level of competence at which they were performing and then to give them feedback aiming to advance them to the next levels of competence.

2.2.2 Internal competential feedback

In the Internal feedback condition students were instructed to self-assess their performance in relation to the above described five levels of scientific reasoning (see section 2.2). They were asked to tick the levels they think they had already thought of in the previous task. They were also told that they could improve in the next task by attending to the requirements indicated by the levels they had not marked.

2.2.3 Control condition

In the No-feedback condition students were given the general information concerning the five levels of scientific reasoning (see section 2.2), but their performance was not evaluated in any way. They were neither told their competence levels by the teacher nor were they instructed to self-assess their performance.

2.3 Measures

2.3.1 Performance

Effects on students' performance in scientific reasoning were investigated by means of three paper-pencil-tasks developed by Mayer et al. (2008), asking students to individually plan scientific experiments. The tasks required the planning of experiments related to the students' everyday life situations and could be planned by the students without any task-specific prior knowledge (e.g., "Tom wants to give some roses to his mother as a present for her birthday. He has read in a gardening journal that roses would grow better in flower soil than in garden mould. Before giving the roses to his mom, he wants to investigate if this

statement in the gardening journal is correct. Please plan an experiment by which the statement can be investigated scientifically. Describe the experiment as accurately as possible“). Directly after task completion, answers were coded by the researcher on a scale from competence levels 1-5 (as noted above) according to a coding rubric provided by Mayer et al. (2008). After study completion, answers were coded again by two blind raters.

2.3.2 Calibration

Two different measures of Calibration were used, based on the research by Schraw et al. (1993). The first measure aimed to assess the direction of self-evaluative judgment errors. This Calibration Bias comprises the difference between true and estimated performance, indicating over- or underconfidence. According to Schraw et al. (1993), the mean bias score was computed by subtracting true performance (i.e., their levels of performance coded by blind raters according to the coding rubric as indicated above) from estimated performance, averaged for all post intervention measuring points. Students estimated their performance before all performance tasks (2.3.1), answering the question how many correct answers they expected themselves to give in the next experiment on a scale from 0 to 5. Mean bias score therefore ranged from -5 to 5, with scores below zero indicating underconfidence, 0 indicating perfect calibration, and scores above zero indicating overconfidence.

The second calibration measure aimed to assess the accuracy of self-evaluative performance judgments. Calibration Accuracy was calculated by the absolute value of the difference between estimated and true judgments, averaged across measuring points. The accuracy score ranged from 0 to 5, with 0 indicating perfect accuracy and 5 indicating a lack of accuracy.

2.3.3 Pre-group differences

Although the students were randomly assigned to the three treatment groups, four measures were used to check whether there were initial differences between the groups. First, a measure of the underlying competence, a nonverbal reasoning scale of the KFT 4-12+R, was used (Heller & Perleth, 2000). The KFT 4-12+R is a German adaptation of the Cognitive Abilities Test developed by Thorndike and Hagen (1971). The *figure analogy* scale contains 25 items selected for 8th grade and these items were presented in a multiple choice, paper-and-pencil format. Second, three items were developed to measure Interest in Biology (as the task related to daily life biology examples; e.g., “I am interested in biology.”). Third, two items were developed to measure Experience in Planning Experiments (e.g., “I have planned experiments like that before”). Fourth, performance on the first task (planning for experiments) was checked to ensure no pre-group differences prior to experimental intervention.

2.4 Procedure

All test instructions were read to the students by the experimenter to achieve comparability between the experimental conditions. Before starting to work on the performance tasks, for which students had a maximum of 10 minutes each, all participants received a copy of the five scientific reasoning steps and each step was explained to them. This was followed by the control measure of Interest in Biology, the first Calibration measure, and the first performance task. Performance of task 1 was then rated by means of a coding manual (Mayer et al., 2008) and the respective feedback sheet was handed out. This was followed by the Calibration measure of task 2 and the planning of the same. No feedback was provided after task 2. As a filler-item, the general Reasoning questionnaire was conducted. Subsequently, the study included a third Calibration measure, planning of task 3

and feedback on it. Finally, the control measure of Experience in Planning Experiments and the demographic items (sex, age, school) concluded the study. The study took 60 minutes. Ethics approval was granted by the Department of Education in Hamburg.

2.5 Methods of analysis

Coefficient alpha was used to estimate reliability, and *MANOVA* to evaluate differences in the dependent variables moderated by group membership. To ensure comparability and remove any differences in the difficulties of the tasks, *z*-standardized means were used in subsequent performance analyses.

3 RESULTS

3.1 Estimates of reliability

The estimates of reliability were calculated for each scale: Performance tasks ($\alpha = .79$, Calibration-Bias $\alpha = .76$; Calibration-Accuracy $\alpha = .76$; Control-Reasoning $\alpha = .85$; Control-Interest in Biology $\alpha = .81$; Control-Experience in Planning Experiments $\alpha = .73$). In all cases the estimates were sufficiently high to provide confidence to use total scores in subsequent analyses.

Inter-rater reliability was used to evaluate the consistency of the coding of the three tasks to the levels. During the experiment coding was undertaken by the researcher as immediate feedback needed to be provided. After the experiment, two researchers independently coded the responses to each task according the rubric in the manual, and these two researchers were blind as to which condition the responses emanated from. The chance-corrected inter-rater reliability of task 1-3 by the two blind raters ranged from $\kappa = 0.74$ to $\kappa = 0.91$ (Cohen, 1960) and therefore was sufficiently high to provide confidence in the consistency of the ratings (Landis & Koch, 1977). Where there were discrepancies, the two

blind raters were requested to agree on a rating. Consistency of this rating with the researchers' ratings during the study after task 1 and after task 3 was also quite high (task 1, $\kappa = 0.89$; task 3, $\kappa = 0.67$; Cohen, 1960). Feedback after task three was provided at the end of the study, hence, only feedback information after task 1 influenced dependent variables. The researchers' ratings after task 1 only differed twice and each with one point from the blind ratings. Thus, the researchers' ratings influencing dependent variables were highly accurate. When correlating students' self-assessments with blind ratings, Cohen's Kappa indicated low consistency ($\kappa = 0.21$). To further analyze this low consistency and to reveal the direction of the evaluation differences (i.e., students' over- or underestimation), students' self-assessments were compared to the blind ratings. A *MANOVA* showed, that students' self-evaluations in task 1 and task 3 differed significantly from the blind ratings ($F(1,48) = 7.23$; $p = .01$; $\eta^2 = .13$): This significant main effect indicates an overestimation of students' self-assessments (SA) when compared to the blind ratings (BR) for both tasks (task 1, $M_{BR} = 2.25$; $M_{SA} = 3.56$; task 3, $M_{BR} = 2.96$; $M_{SA} = 3.36$). This fact will be further discussed (see section 4.1).

The ratings the two blind raters had agreed on entered main data analyses.

3.2 Control variables

A *MANOVA* did not show any statistically significant differences across the three feedback groups in the mean scores for general reasoning, experience in planning experiments and interest in biology (*Wilks Lambda* = .93; *Mult.F*(6,128) = 0.79; $p = .59$), nor were there any univariate differences for each of the three variables (General Reasoning, $F(2,66) = 0.30$; $p = .74$; Interest in Biology, $F(2,66) = 1.84$; $p = .17$; Planning Experiments, $F(2,66) = 0.70$; $p = .50$). Neither differed the experimental groups in their performance on

task 1 ($F(2,72) = 0.89$; $p = .42$). Thus, there were not any pre-group differences prior to experimental intervention.

3.3 Main analyses

The main analyses concerned the impact of feedback variation on performance and calibration. Means and standard deviations of the following analyses are all shown in Table 1.

Table 1

Means (standard deviations) concerning the dependent variables Performance (z-standardized) and Calibration, i.e. Bias and Accuracy of self-evaluative performance judgments

	Competential Feedback		Control group
	external ($n = 25$)	internal ($n = 25$)	no-feedback ($n = 25$)
Performance			
Task 1 (Pretest)	0.21 (1.06)	-0.04 (0.92)	-0.16 (1.02)
Task 2 (Posttest)	0.42 (0.97)	-0.26 (0.96)	-0.16 (0.97)
Task 3 (Follow-up)	0.57 (0.94)	-0.27 (1.08)	-0.30 (0.71)
Calibration			
Bias (Pretest)	0.40 (1.50)	0.92 (1.35)	1.40 (1.47)
Bias (Posttest)	-0.06 (1.24)	1.10 (1.23)	1.46 (1.12)
Accuracy (Pretest)	1.12 (1.05)	1.32 (0.95)	1.56 (1.29)
Accuracy (Posttest)	1.18 (0.84)	1.62 (0.78)	1.66 (0.92)

Note: Performance was coded on competence levels 1-5; a higher number indicating a higher competence level. Calibration bias is the difference between true and estimated performance indicating over- or underconfidence; mean bias score ranged from -5 to 5, with 0 indicating perfect calibration. Calibration accuracy is the absolute value of the difference between true and estimated performance; the accuracy score ranged from 0 to 5, with 0 indicating perfect accuracy.

3.3.1 Performance

Participants' performance in scientific reasoning was examined by means of three tasks (Mayer et al., 2008). The first measuring point was set prior to the experimental variation (i.e., feedback variation); it thus can be considered baseline information or pretest performance. Measuring point 2 comprised the posttest measuring point after the experimentally varied feedback had been provided. As a follow-up test, measuring point three examined performance without previous feedback information. Students did not differ in their pretest performance (see 3.2). A repeated measures *MANOVA* with the factors *time* (t2-t3) and *feedback* was used to analyze feedback effects on performance. There was not a significant interaction between the factors *time* and *feedback* ($F(2,72) = 0.67$; $p = .51$). More importantly, the main effect for *feedback* ($F(2,72) = 6.73$; $p < .01$; $\eta^2 = .16$) was highly significant. Hence, for both measuring points, Bonferroni's post-hoc test revealed that students in the External feedback condition performed significantly better than students in the Internal feedback or No-feedback control condition; there was not a significant difference between the Internal and the No-feedback condition (see Table 1). *Z*-standardized means (*t*-transformed) of students' performance in the posttest and follow-up-measuring point are shown in Figure 1.

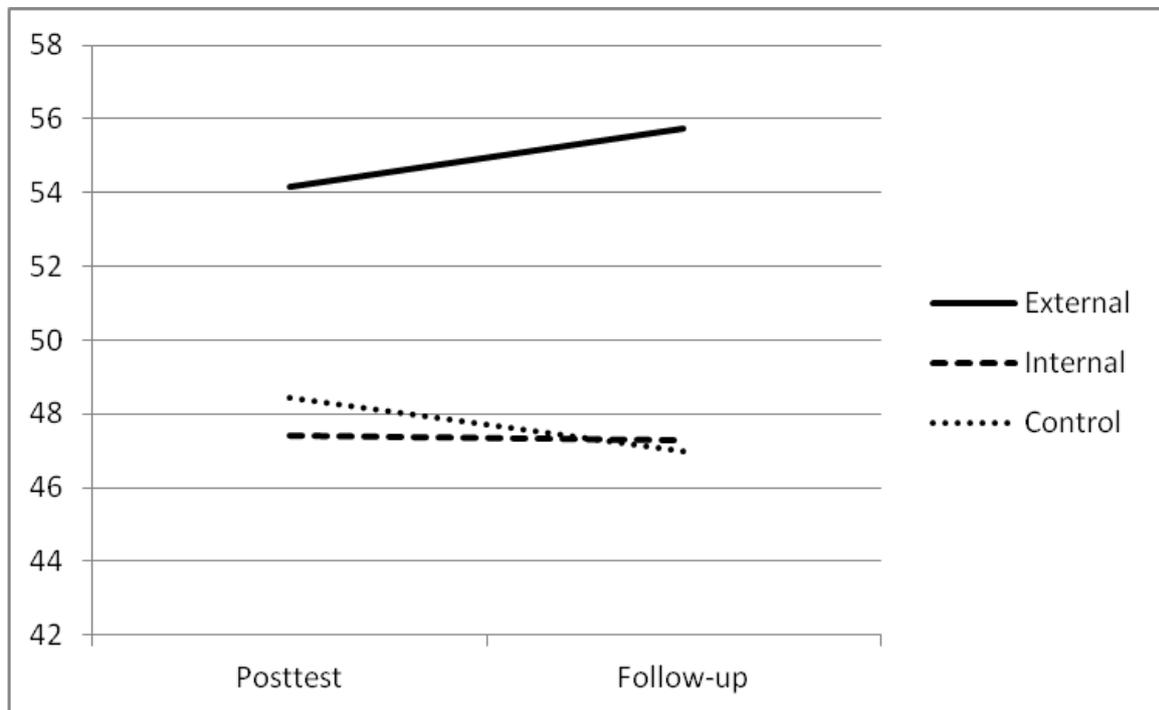


Figure 1. *T*-transformed means of the *z*-standardized means of students' performance in the posttest and follow-up-measuring point dependent on the competential feedback source (i.e., external teacher feedback vs. internal students' self-assessment) vs. no feedback information (i.e., control).

3.3.2 Calibration

Calibration (i.e., Bias and Accuracy of self-evaluative performance judgments) was measured with three items; the first item was set prior to the experimental variation. Concerning this pretest, intervention conditions did not differ in both calibration measures (Bias, $F(2,72) = 3.01$; $p = .06$; Accuracy, $F(2,72) = 0.99$; $p = .38$).

In the Calibration posttest (i.e., the remaining two items), a *MANOVA* showed that there were differences across the three groups (*Wilks Lambda* = .76; $\text{Mult.}F(4,142) = 5.25$; $p < .001$; $\eta^2 = .13$), and that both Bias ($F(2,72) = 11.07$; $p < .001$; $\eta^2 = .24$) and Accuracy ($F(2,72) = 2.46$; $p = .04$; $\eta^2 = .06$) contributed to this overall difference. Although participants

of all intervention groups were overconfident, planned contrasts revealed that there was stronger Bias in the Internal feedback ($t(2,72) = 3.44; p < .001; d = .85$) and the Control condition ($t(2,72) = 4.50; p < .001; d = 1.19$) compared to the External feedback condition. Mean bias did not differ significantly between the internal feedback vs. the control condition ($t(2,72) = 1.07; p = .15; d = .31$).

With regard to Calibration Accuracy, self-evaluative performance judgments in the External feedback condition were significantly more accurate than in the Internal feedback condition ($t(2,72) = 1.83; p = .04; d = .54$) and in the Control condition ($t(2,72) = 2.00; p = .02; d = .54$). Calibration Accuracy did not differ significantly between the Internal feedback vs. the Control condition ($t(2,72) = 0.17; p = .44, d = .05$).

3.4 Mediation analyses

The next analysis investigated a potential underlying variable of the differential effects of External and Internal feedback on performance. On the one hand, it included linear regression analyses with regard to the impact of Calibration Accuracy on performance. On the other hand, it was examined whether Calibration Accuracy mediated the positive impact of External feedback on performance. Therefore, regression analyses were performed. Performance was the mean of z -standardized participants' posttests and follow-up performance (t_2-t_3). The Calibration Accuracy coefficient as the mean of the absolute value difference between estimated and true performance judgments entered the analyses. To be consistent in the direction of the effects, this coefficient was recoded before entering the analyses so that a higher number indicated higher accuracy. The feedback coefficient comprised the contrast between the external (2) and the internal (-1) feedback, and the control (-1) condition.

First, the independent factor “feedback” had a significant effect on the dependent variable “performance”, $\beta = .37$, $t(73) = 3.40$, $p = .001$. The experimentally varied feedback information explained a significant proportion of variance of performance, $R^2 = .14$, $F(1,73) = 11.57$, $p = .001$. Second, according to the regression analyses, the independent factor “feedback” was a significant predictor of the mediator variable “calibration accuracy”, $\beta = .25$, $t(73) = 2.23$, $p = .03$. The experimentally varied feedback information explained a significant proportion of variance of calibration accuracy, $R^2 = .06$, $F(1,73) = 4.96$, $p = .03$. Third, concerning the impact of Calibration Accuracy on performance, the accuracy of self-evaluative performance judgments predicted the dependent variable “performance”, $\beta = .46$, $t(73) = 4.45$; $p < .001$. Calibration Accuracy explained a significant proportion of variance of performance, $R^2 = .21$, $F(1,73) = 19.82$, $p < .001$.

Therefore, Calibration Accuracy can be a mediator for the impact of External feedback on performance as the present study showed. This is due to the fact that in all three regression analyses, the predictor significantly predicts the criterion which is needed to be met to prove mediation (Baron & Kenny, 1986).

When regressing “feedback” and “calibration accuracy” as predictors of “performance”, the effect of the independent variable decreased ($\beta = .27$), but still remained significant ($t(72) = 2.62$; $p = .01$). This indicates partial mediation.

The Sobel test (Sobel, 1982) was used to investigate whether the partial mediator variable “calibration accuracy” significantly conveys the influence of feedback on performance. It is a significance test of indirect impact (see MacKinnon, Warsi, & Dwyer, 1995). The Sobel test reached statistical significance indicating that the mediation was statistically significant (SEb *Calibration accuracy Feedback.Performance* Calibration accuracy = .16; b *Calibration accuracy Feedback.Performance* Calibration accuracy = .62; $t_{\text{Sobel}} = 1.88$; $p = .03$).

4 DISCUSSION

4.1 Conclusions

This study aimed to investigate competential feedback effects on learning outcomes dependent on the feedback source (i.e., external or internal feedback). The competence model was used as an external standard to which students' achievements were compared to and either fed back to them by the teacher (i.e., the researcher) or self-assessed by the students themselves.

Previous studies on external competential feedback found a beneficial effect on students' performance and self-evaluative performance judgments (Wollenschläger et al., 2011a). *First*, this study replicated the beneficial performance effect in a setting when there was more time between feedback and performance task: In the follow-up test, students with external competential feedback performed better than students with internal feedback or without any feedback, even when the feedback information was not provided before task completion. Thus, the beneficial feedback effect on performance was shown without being confounded with prompting effects indicating a strong and persistent feedback effect.

Second, competential feedback was assumed to be effective also in self-assessment since it informs students about their correct answers (how am I going?), it is task-specific, elucidates goals and includes cues on the current state of their learning process (where am I going?) as well as on how to improve their task performance (where to next?). These feedback aspects have been identified as critical in making feedback powerful (e.g., Hattie, 2009). Thus, it was assumed that both, external as well as internal feedback, improve students' task performance and self-evaluative performance judgments compared to a control condition without feedback information. Contrary to our hypotheses, only external feedback significantly improved students' performance and calibration accuracy: Students in the external feedback condition showed a significantly better task performance in planning

scientific experiments and had less biased and more accurate self-evaluative performance judgements when compared to the internal feedback and the control condition. Students self-assessing their performance relative to the competency levels did not differ from the control group without any feedback information in both dependent variables (i.e., performance and calibration). This finding is striking: When using the same assessment criteria, external and internal students' feedback have differential effects on students' performance and calibration. This highlights the importance of mediating variables discovering the underlying mechanisms of feedback effectiveness. Reported findings concerning students overestimating their self-evaluations point to the fact that accuracy of feedback seems crucial to make feedback powerful. This hypothesis is supported in the following analyses:

With regard to our third hypothesis, calibration accuracy has been identified as a mediator, although only for external competential feedback effects on performance: teachers' feedback given relative to competency levels induced more accurate self-evaluative performance judgments which in turn improved students' task performance. Students self-assessing their performance relative to competency levels seem to remain overconfident (just as students without feedback information) which consequently impedes their performance. This underlines the importance of studying how feedback information is received and processed on a metacognitive level rather than how it is provided (Hattie & Wollenschläger, 2012).

4.2 Limitations and future studies

These results have to be interpreted with regard to the following limitations: this study investigated feedback effects in a short intervention time using three performance tasks of one domain (i.e., planning of experiments). To a) investigate the development of feedback effects over time and b) to be able to generalize the findings to other domains, further studies

should investigate feedback effects over a longer period of time using more and diverse performance tasks.

Also, this study's findings on competential self-assessment can only be generalized to a limited extent: self-assessment relative to competency levels was not trained before it was compared to teacher feedback. Self-assessing one's own performance involves a complex process of internalization (Andrade & Du, 2007) and self-assessment-training seems one crucial factor to enhance performance (Ross, 2006). Thus, competential internal feedback effects should be further investigated together with training students on how to assess their own performance (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009; Kasanen & Rätty, 2002) also to improve accuracy of self-evaluations. In this context, future studies could also transfer the study into the school setting to improve the ecological validity of the findings. By doing so, one could also focus on investigating a combination of external and internal competential feedback which would be applicable to classroom assessment. However, as soon as the teacher variable is not kept constant, teachers' judgment accuracy, for example their level component (Südkamp, Kaiser, & Möller, in press) also needs to be considered.

To clearly explain differential external vs. internal feedback effects on performance, other mediators or moderators need identification. Possible mediator or moderator effects on the way how feedback is received and whether learning is enhanced could be perceived self-efficacy (Cleary & Zimmerman, 2004; Kitsantas & Zimmerman, 2006), students' pre-conditions such as high-achieving vs. low-achieving students (Chappuis & Stiggins, 2002) or students' under- vs. overconfidence (Black & Wiliam; 1998; Labuhn et al., 2010).

Another promising approach for future studies is linking competential feedback with the research done in peer feedback (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010). In peer feedback research, feedback quality in terms of accuracy and/or concordance with teacher feedback (Strijbos, Narciss, & Dünnebier, 2010; Van Steendam, Rijlaarsdam,

Sercu, & Van den Bergh, 2010) has been identified as central in improving students' performance. These feedback attributes could be transferred to and further studied in self-assessment, too.

4.3 Implications

This study underlines the effectiveness of competential teachers' feedback. Also, the importance of accurate self-evaluations in self-assessment and feedback reception is stressed. Implications can be drawn for educational research and practice:

The theoretical contribution of this study shows that effective teacher feedback does not automatically transfer to self-assessment feedback. Research needs to be done on the training of competential self-assessment and on combining it with external competential feedback or (competential) peer-feedback.

Regarding the educational implications, competential teacher feedback seems to be a promising approach for feedback in schools. Beyond that, whenever applying feedback in schools (i.e., independent of the feedback source), calibration accuracy needs to be enhanced: accurate students' performance judgments are crucial in making feedback powerful.

5 REFERENCES

- Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for assessment in New-Zealand: Developing students' assessment capabilities*. Wellington, NZ: Ministry of Education.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 159-181. doi:10.1080/02602930600801928
- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27, 3-13. doi:10.1111/j.1745-3992.2008.00118.x
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48, 12-19. doi:10.1080/00405840802577544
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238. doi:10.3102/00346543061002213
- Baron, K. E., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration, *Journal of Personality and Social Psychology*, 51, 1173-1182. doi:10.1037/0022-3514.51.6.1173
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, 18, 5-25. doi:10.1080/0969594X.2010.513678
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74. doi:10.1080/0969595980050102

- Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*, 269-290. doi:10.3200/JEXE.73.4.269-290
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529-549. doi:10.1007/BF00138746
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*, 399-413. doi:10.1080/02602930600679050
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research, 65*, 245-281. doi:10.3102/00346543065003245
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*, 40-43.
- Chen, P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 79-92. doi:10.1016/j.lindif.2003.08.003
- Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A schoolbased program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools, 41*, 537-550. doi:10.1002/pits.10177
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. doi:10.1177/001316446002000104
- Connell, M. W., Sheridan, K., & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126-155). Cambridge, MA: Cambridge University Press.

- Cooper, D. (2006). *Talk about assessment: Strategies and tools to improve learning*. Toronto, ON: Thomson Nelson. Government of British Columbia.
- Cyboran, V. (2006). Self-assessment: Grading or knowing? *Academic Exchange Quarterly*, 10, 183-186.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331-350. doi:10.1080/03075079912331379935
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395-430. doi:10.3102/00346543059004395
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20, 304-315. doi:10.1016/j.learninstruc.2009.08.007
- Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97, 21-34. doi:10.1080/00220670309596625
- Harks, B., Rakoczy, K., Hattie, J., Klieme E., & Besser, M. (2011). *Indirect feedback effects on mathematics achievement*. Manuscript submitted for publication.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi:10.3102/003465430298487
- Hattie, J., & Wollenschläger, M. (2012). *A conceptualization of feedback*. Manuscript submitted for publication.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeits-Test für 4.-12. Klassen, Revision (KFT 4-12+R)* [Cognitive Ability Test for grades 4 to 12, revised version]. Göttingen: Beltz.
- Hoffman, B., & Spatariu, A. (2008). The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology*, 33, 875-893. doi:10.1016/j.cedpsych.2007.07.002
- Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, 20, 18-29. doi:10.1016/j.learninstruc.2008.12.001
- Kasanen, K., & Rätty, H. (2002). "You be sure now to be honest in your assessment": Teaching and learning self-assessment. *Social Psychology of Education*, 5, 313-328. doi:10.1023/A:1020993427849
- Kitsantas, A., & Zimmerman, B. J. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning*, 1, 202-212. doi:10.1007/s11409-006-9000-7
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht [Performance assessment and competence modeling in mathematics instruction]. *Zeitschrift für Pädagogik*, 56, 64-74.

- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3-22). Göttingen: Hogrefe.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen [Competence models for assessing individual learning outcomes and evaluating educational outcomes]. *Zeitschrift für Pädagogik*, 52, 876-903. Retrieved from http://www.pedocs.de/frontdoor.php?source_opus=4493
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284. doi:10.1037//0033-2909.119.2.254
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Journal of Psychology*, 216, 61-73. doi:10.1027/0044-3409.216.2.61
- Kruger, J., & Dunning, D. (2009). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134. doi: 10.1037/0022-3514.77.6.1121
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232. doi:10.3102/00346543047002211
- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173-194. doi:10.1007/s11409-010-9056-2

- Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35*, 135-156. doi:10.1080/02602930802687737
- Landis, G., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62. doi:10.1207/s15327906mbr3001_3
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung [Competence model of Scientific Inquiry]. In U. Harms, & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* [Research in Teaching and Instruction in Biology Education] (pp. 63-79). Innsbruck: Studienverlag.
- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.
- Nolen, S. B. (1996). Why study? How reasons for learning influence strategy selection. *Educational Psychology Review, 8*, 335-355. doi:10.1007/BF01463938
- Nulty, D. D. (2011). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education, 36*, 493-507. doi:10.1080/02602930903540983
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation, 11*, 1-13.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment, 8*, 43-59. doi:10.1207/S15326977EA0801_03

- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing*, 6, 107-132. doi:10.1016/S1075-2935(99)00003-3
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. doi:10.1007/BF00117714
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5, 77-85. doi:10.1080/0969595980050104
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, 18, 455-463. doi:10.1006/ceps.1993.1034
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189. doi:10.3102/0034654307313795
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290-312). Washington: American Sociological Association.
- Stiggins, R., & Chapuis, J. (2006). What a difference a word makes: Assessment FOR learning rather than assessment OF learning helps students succeed. *Journal of Staff Development*, 27, 10-14.
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12, 437-475. doi:10.40-726X/00/1200-0437
- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20, 291-303. doi:10.1016/j.learninstruc.2009.08.008

- Südkamp, A., Kaiser, J., & Möller, J. (in press). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*.
- Thorndike, R. L., & Hagen, E. P. (1971). *Cognitive Abilities Test*. Boston: Houghton Mifflin.
- Towler, L., & Broadfoot, P. (1992). Self-assessment in primary school. *Educational Review*, 44, 137-151. doi:10.1080/0013191920440203
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20, 316-327. doi:10.1016/j.learninstruc.2009.08.009
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe & Huber.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 18-31. doi:10.1207/s1532690xci1601_2
- Wollenschläger, M., Hattie, J., Möller, J., & Harms, U. (2011a). *Competential feedback effects on performance: are they mediated by calibration?* Manuscript submitted for publication.
- Wollenschläger, M., Möller, J. & Harms, U. (2011b). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken [Effects of competential feedback on scientific reasoning]. *Zeitschrift für Pädagogische Psychologie*, 25, 197-202. doi:10.1024/1010-0652/a000040

ZUSAMMENFASSUNG

EXTERNER VERSUS INTERNER KOMPETENZIELLES FEEDBACK: ZEIGEN SICH DIFFERENZIELLE EFFEKTE AUF PERFORMANZ UND GENAUIGKEIT DER SELBSTEINSCHÄTZUNG?

In der Literatur werden externe Feedback-Quellen (z.B. Fremdbewertung eigener Leistung) und interne Feedback-Quellen (d.h. Selbstbewertung eigener Leistung) unterschieden (z.B. Butler & Winne, 1995). Für beide Feedbackquellen sind positive Effekte auf den Lernprozess nachgewiesen (vgl., Hattie, 2009; Hattie & Wollenschläger, 2012). Die vorliegende Studie untersucht Effekte der Feedbackquelle (Fremd- vs. Selbstfeedback) im Bereich des kompetenziellen Feedbacks und zwar auf Performanz und der diesbezüglichen Genauigkeit der Schülereinschätzungen.

Kompetenzielles Feedback wird als Leistungsbewertung in Relation zu Stufen eines domänenspezifischen Kompetenzmodells verstanden. Bisherige Studien zu Effekten dieses kriterialen Feedbacks auf Performanz und Genauigkeit der Schülerelbsteinschätzung vergleichen auf Kompetenzstufen basierendes Fremdfeedback im Vergleich zu keiner Feedbackinformation oder im Vergleich zu sozialer Feedbackinformation (Einstufung der Schülerleistungen in Relation zum Klassendurchschnitt, vgl. Wollenschläger, Hattie, Möller, & Harms, 2011a). Ergebnisse dieser Studien zeigen, dass Lernende mit kompetenziellem Fremdfeedback signifikant besser abschneiden und ihre eigene Leistung signifikant genauer einschätzen als Lernende mit sozialem oder keinem Feedback; die Genauigkeit der Performanzeinschätzung vermittelt den positiven Effekt von kompetenziellem Fremdfeedback auf die Performanz der Lernenden. Effekte von kompetenziellem Selbstfeedback, d.h. die Einstufung der eigenen Performanz in Relation zu Kompetenzstufen, wurden bisher noch nicht untersucht.

Bisherige Studien zur Selbstbewertung der eigenen Leistung weisen sowohl positive Effekte auf Performanz (vgl. Andrade, Du, & Wang, 2008) als auch auf metakognitive

Prozesse von Lernenden nach (vgl. Cooper, 2006). Auch Selbstbewertungen der eigenen Leistung anhand vorgegebener Rubriken sind bereits untersucht (Andrade & Valtcheva, 2009). Die vorgegebenen Rubriken enthalten die zu erreichenden Lerninhalte; durch die Selbstbewertung der eigenen Leistung in Relation zu den Rubriken können die Lernenden ihre erreichten und noch zu erreichenden Lerninhalte evaluieren. Diese sogenannten *standards-referenced assessment tools* (Goodrich Andrade & Boulay, 2003) sind bisher a) nur im Vergleich zu einer Kontrollgruppe ohne Selbsteinschätzungsintervention und b) ausschließlich hinsichtlich potenziell positiver Performanzeffekte untersucht worden. Eine Untersuchung der vergleichenden Effekte einer vorgegebenen Rubrik, die sowohl für Fremd- als auch für Selbstfeedback genutzt wird und deren Effekte auf Performanz wie auch metakognitive Variablen untersucht werden, steht noch aus. Die vorliegende Studie greift diesen Forschungsbedarf im Kontext des kompetenziellen Feedbacks auf: Erstens wird kompetenzielles Fremdfeedback mit kompetenziellem Selbstfeedback (Einstufung der eigenen Performanz in Relation zu Kompetenzstufen) verglichen. Zweitens werden Effekte auf die Performanz der Lernenden wie auch metakognitive Variablen (in diesem Fall die Genauigkeit der Schülerelbsteinschätzung) untersucht.

In einem Pretest-Posttest-Follow-up-Design wurden Schülerinnen und Schüler der Sekundarstufe 1 ($N = 75$) entweder einer Bedingung mit externem kompetenziellem Feedback (Fremdfeedback) oder einer internen Feedbackbedingung (Selbstbewertung in Relation zu den Kompetenzstufen) oder einer Kontrollgruppe ohne Feedback randomisiert zugewiesen.

Lernende mit externem Feedback schnitten signifikant besser ab und schätzten ihre eigene Leistung signifikant genauer ein als Lernende mit internem oder keinem Feedback. Die Genauigkeit der Performanzeinschätzung konnte als partieller Mediator des Feedbackeffekts auf Performanz identifiziert werden: kompetenzielles Fremdfeedback führte

zu genaueren Selbsteinschätzungen der Lernenden, diese wiederum trugen zu einer Performanzsteigerung bei.

Somit unterstreicht die Studie erstens den Nutzen kompetenziellen Fremdfeedbacks und zweitens die zentrale Wirkung von Selbsteinschätzungen in der Wirkung von Feedbackinformationen.

STUDIE 4

IST KOMPETENZIELLES FREMDFEEDBACK ÜBERLEGEN,
WEIL ES ALS EFFEKTIVER WAHRGENOMMEN WIRD?

Wollenschläger, M., Möller, J., & Harms, U. (2011). *Ist kompetenzielles
Fremdfeedback überlegen, weil es als effektiver wahrgenommen wird?*

Manuskript eingereicht zur Publikation.

ZUSAMMENFASSUNG

Kompetenzielles Feedback ist als Leistungsbewertung in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells definiert. Diese Studie untersucht differenzielle Effekte dieser Form des kriterialen Feedbacks auf Performanz und wahrgenommene Feedbackeffektivität in Abhängigkeit von der Feedbackquelle. In einem Pre-Posttest-Design wurden Schülerinnen und Schüler der Sekundarstufe 1 ($N = 50$) entweder einer Bedingung mit Fremdbewertung ihrer Leistung in Relation zu Kompetenzstufen (externes kompetenzielles Feedback) oder mit Selbstbewertung ihrer Leistung in Relation zu Kompetenzstufen (internes kompetenzielles Feedback) zugewiesen. Lernende mit externem Feedback schnitten signifikant besser ab und nahmen das Feedback signifikant als effektiver wahr als Lernende mit internem Feedback. Die wahrgenommene Feedbackeffektivität medierte die differenziellen Feedbackeffekte auf Performanz vollständig. Die Studie unterstreicht a) die Effektivität kompetenziellen Fremdfeedbacks und b) die zentrale Rolle der wahrgenommenen Effektivität in der Feedbackverarbeitung.

ABSTRACT

Competential feedback is defined as the evaluation of students' achievements in relation to the levels within a domain-specific competence model. This study aimed at investigating differential effects of this criterion-related feedback on performance and perceived feedback effectiveness depending on the source of feedback information. In a pre-post experimental design, randomly assigned secondary school students ($N = 50$) either received external competential feedback (i.e., outside evaluation relative to the competency levels) or internal competential feedback (i.e., self-assessed their performance relative to the competency levels). Students in the external feedback condition showed a significantly better task performance in planning scientific experiments and perceived their feedback information significantly more effective than those students in the internal feedback condition. This perceived feedback effectiveness totally mediated the differential feedback effects on performance. This study underlines a) the benefit of external competential feedback and b) the importance of how feedback effectiveness is perceived in making feedback powerful.

1 EINLEITUNG

In der Feedbackliteratur werden externe Quellen (z.B. Fremdbewertung eigener Leistung) und interne Quellen (d.h. Selbstbewertung eigener Leistung) von Feedbackinformationen unterschieden (z.B. Butler & Winne, 1995). In speziellen Varianten externen und internen Feedbacks wird von kompetenziellem Feedback gesprochen (Wollenschläger, Möller, & Harms, 2012). Kompetenzielles Feedback ist definiert als Leistungsbewertung in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells; die Lernenden erhalten nach der Aufgabenbearbeitung Informationen zu den von ihnen erreichten und/oder noch zu erreichenden Kompetenzstufen (Wollenschläger, Möller, & Harms, 2011). Dieses kompetenzielle Feedback ist sozial vergleichendem Feedback überlegen (Wollenschläger, Hattie, Möller, & Harms, 2011). Bisherige Befunde zum Vergleich zwischen externem und internem kompetenziellem Feedback (Wollenschläger et al., 2012) weisen einen Überlegenheitseffekt des Fremdfeedbacks nach: Lernende, deren Performanz von der Lehrkraft in Relation zu den erreichten und noch nicht erreichten Kompetenzstufen eingeschätzt und rückgemeldet wird, schneiden signifikant besser ab und schätzen ihre Performanz anschließend signifikant genauer ein als Lernende, die ihre Performanz in Relation zu den Kompetenzstufen selbst bewerten. Zentrale Forschungsfrage der vorliegenden Studie ist, inwieweit die wahrgenommene Feedbackeffektivität der Überlegenheit des kompetenziellen Fremdfeedbacks zugrundeliegt.

1.1 Feedbackeffektivität

Feedback ist definiert als Information, die zeitlich während oder nach einem Prozess oder Prozessschritt angesiedelt ist, um regulierend auf Prozess- oder Prozessschritt-Ebene einzuwirken (vgl. Hattie & Gan, 2011; Narciss, 2008). Lehrerfeedback im Unterricht wurde mit einer Effektstärke von $d = 0.79$ (basierend auf 12 Meta-Analysen, 196 Studien und

6972 Effektstärken; Hattie, 2009) als essenzielle Einflussgröße effektiver Lehr-/Lernprozesse identifiziert. Jedoch weisen die in den Meta-Analysen berichteten Effektstärken eine hohe Variation auf, und nicht jede Form des Feedbacks scheint sich positiv auf den Lernprozess auszuwirken (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie, 2009; Kulhavy, 1977). Meta-Analysen besagen zusammenfassend, dass Feedback das spezifische Lernziel sowie den erreichten Lernstand in Relation zu diesem Lernziel verdeutlichen und vor allem Informationen bereithalten sollte, wie die Performanz zukünftig verbessert werden kann (Black & Wiliam, 1998; Hattie, 2009, 2012; Hattie & Gan, 2011; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mory, 2004; Sadler, 1989, 1998; Shute, 2008). Hattie und Gan (2011, basierend auf Hattie & Timperley, 2007, p. 87) fassen die Kriterien effektiven Feedbacks in ihrem *visible learning and teaching model* in drei Fragen zusammen: Effektives Feedback beantwortet demnach die Fragen „*Where am I going*“ (Was ist das Lernziel?), „*How am I going*“ (Wie hat der oder die Lernende in Relation zum Lernziel abgeschnitten?) und „*Where to next*“ (Wie kann das Lernziel zukünftig besser erreicht werden?). Es wird angenommen, dass kompetenzielles Feedback diese Kriterien effektiven Feedbacks beinhaltet (Wollenschläger et al., 2011a).

1.2 Kompetenzielles Feedback

Im Kontext der zunehmenden Fokussierung auf Bildungsziele und deren Konkretisierung in den Bildungsstandards hat die Entwicklung von Kompetenzmodellen in der pädagogisch-psychologischen und fachdidaktischen Forschung einen zentralen Platz eingenommen (z.B. Klieme, Bürgermeister, Harks, Blum, Leiß, & Rakoczy, 2010; Köller & Parchmann, 2012; Viering, Fischer, & Neumann, 2010). Kompetenzmodelle bilden die Basis nationaler und internationaler Schulleistungsvergleichsstudien im Rahmen summativer Evaluationen auf der Bildungssystemebene; sie ermöglichen jedoch auch auf Individualebene

(Schülerinnen und Schüler) den individuellen Leistungsstand und die Leistungsentwicklung abzubilden und formativ zurückzumelden (Bernholt, Parchmann, & Commons, 2009; Koeppen, Hartig, Klieme, & Leutner, 2008).

Kompetenzmodelle zielen darauf ab, die Struktur (Dimensionen oder Teilkompetenzen) einer domänenspezifischen Kompetenz darzustellen und beinhalten, basierend auf Theorien und empirischen Befunden, definierte Anforderungsbereiche derselben (Graduierung in Kompetenzstufen nach zunehmender Komplexität, siehe Bernholt et al., 2009; Klieme et al., 2003; Schecker & Parchmann, 2006). Die theoretisch abgeleiteten, hierarchisch angeordneten Kompetenzstufen bieten „eine Alternative zur willkürlichen Setzung von Leistungsmerkmalen auf einem Kontinuum“ (Klieme et al., 2003, S. 81) und entsprechen einem objektiv festgelegten und theoretisch begründeten Standard an Lernzielen (Jäger, 2008).

Wird die Performanz eines Lernenden auf Individualebene in Relation zu den Stufen eines Kompetenzmodells bewertet, so wird dies als kompetenzielles Feedback definiert (Wollenschläger et al., 2011a). Im kompetenziellen Feedback werden die Kompetenzstufen als domänenspezifisches Lernziel veranschaulicht („*Where am I going?*“), die erreichten Kompetenzstufen geben Auskunft über den erreichten Lernstand in Relation zum Lernziel („*How am I going?*“), und die (noch) nicht erreichten Kompetenzstufen halten Informationen bereit, wie die Performanz zukünftig verbessert werden kann („*Where to next?*“). Folglich kann kompetenzielles Feedback alle Kriterien effektiven Feedbacks (siehe 1.1) beinhalten und die von Hattie und Timperley (2007) genannten Feedbackfragen beantworten.

1.2.1 Externes kompetenzielles Feedback

Externes kompetenzielles Feedback ist definiert als Fremdbewertung einer Leistung in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells (Wollenschläger et

al., 2012). In bisherigen Studien zu den Effekten externen kompetenziellen Feedbacks wurde ein Kompetenzmodell aus der Domäne wissenschaftliches Denken im naturwissenschaftlichen Unterricht (Teilkompetenz „Planung von Experimenten“) genutzt. Die Rückmeldung gab jeweils eine Versuchsleiterin. Die Studien zeigten, dass externes kompetenzielles Feedback zu einer Performanzsteigerung bei der Planung von Experimenten führt und zwar sowohl im Vergleich zu einer Kontrollgruppe ohne Feedback (Wollenschläger et al., 2011b, 2012) als auch im Vergleich zu sozial vergleichendem Feedback (Einstufung der Schülerleistungen in Relation zum Klassendurchschnitt, siehe Wollenschläger et al., 2011a). In Wollenschläger et al. (2012) wurde bereits ein Vorteil des kompetenziellen Fremdfeedbacks gegenüber einem internen kompetenziellen Feedback (d.h. Selbstbewertung der eigenen Leistung in Relation zu Kompetenzstufen) gefunden.

Die vorliegende Studie erweitert die bisherigen Befunde zu externem kompetenziellem Feedback, indem zusätzlich zu Performanzeffekten auch Effekte auf die wahrgenommene Feedbackeffektivität analysiert werden.

1.2.2 Internes kompetenzielles Feedback

Internes kompetenzielles Feedback ist als Selbstbewertung der eigenen Leistung in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells definiert (Wollenschläger et al., 2012). Der Prozess der Selbstbewertung der eigenen Leistung wird in der Literatur zu formativem Assessment als zentraler Wirkfaktor der Performanzsteigerung durch Feedback benannt (z.B. Bennett, 2011; Black & Wiliam, 1998; Sadler, 1998). So wird beispielsweise angenommen, dass Lernende sich selbst lernförderliches Feedback geben können, wenn sie dafür angemessene Bedingungen und Unterstützung erhalten (Goodrich Andrade & Boulay, 2003). Basierend auf dieser Hypothese untersuchten Goodrich Andrade und Boulay (2003) Effekte von Selbstbewertungen der eigenen Leistung anhand

vorgegebener Rubriken. Die vorgegebenen Rubriken enthielten die zu erreichenden Lerninhalte. Durch die Selbstbewertung der eigenen Leistung in Relation zu den Rubriken konnten die Lernenden ihre erreichten und noch zu erreichenden Lerninhalte evaluieren. Diese sogenannten *standards-referenced assessment tools* verbesserten die Performanz von Lernenden im Bereich der Schreibleistung im Vergleich zu einer Kontrollgruppe ohne Selbstbewertungsintervention (vgl. Goodrich Andrade & Boulay, 2003). Die positiven Effekte des Selbstfeedbacks zeigten sich bei kompetenziellem Feedback nicht (Wollenschläger et al., 2012): Lernende mit Selbstfeedback in Relation zu Kompetenzstufen unterschieden sich in ihrer Performanz nicht von den Lernenden aus der Kontrollbedingung ohne Feedback und schnitten bei der Planung eines Experimentes im Vergleich zu Lernenden mit kompetenziellem Fremdfeedback signifikant schlechter ab. Die vorliegende Studie fokussiert auf diese unterschiedlichen Befunde zwischen externem und internem Feedback. Dabei wird die wahrgenommene Feedbackeffektivität als Erklärungsmöglichkeit der unterschiedlichen Feedbackeffekte in Abhängigkeit der Feedbackquelle (extern versus intern) untersucht.

1.3 Mediatoren differenzieller Feedbackeffekte

Da Feedbackeffekte von Motiven und Zielen der Lernenden abhängen (Butler & Winne, 1995; Nolen, 1996), beziehen aktuelle Forschungsvorhaben vermittelnde Variablen des positiven Feedbackeffekts auf Performanz ein (z.B. Harks, Rakoczy, Hattie, Klieme, & Besser, 2011; Wollenschläger et al., 2011a). Dieser Ansatz wird in der vorliegenden Studie zur Analyse der Feedbackeffekte in Abhängigkeit von der Feedbackquelle (extern vs. intern) genutzt. In der Untersuchung von externem versus internem kompetenziellem Feedback (Wollenschläger et al., 2012) konnte die Genauigkeit der Performanzeinschätzung (*calibration accuracy*, vgl. Schraw, Potenza, & Nebelsick-Gullet, 1993) als partieller

Mediator des Feedbackeffekts auf Performanz identifiziert werden: ausschließlich kompetenzielles Fremdfedback führte zu genaueren Selbsteinschätzungen der Lernenden, diese wiederum trugen zu einer Performanzsteigerung bei. Da es sich um einen partiellen Mediator handelte, liegt es nahe, weitere potenzielle Mediatoren der divergierenden Effekte von externem versus internem kompetenziellem Feedback auf die Performanz der Lernenden zu untersuchen. Anknüpfend an bestehende Literatur soll in der vorliegenden Studie ein mediierender Einfluss der wahrgenommenen Feedbackeffektivität untersucht werden. Im Bereich des prozessorientierten Feedbacks (Harks et al., 2011) wurde die Wahrnehmung der Nützlichkeit des Feedbacks als vermittelnde Variable der Feedbackeffekte auf Performanz (Lösen von Mathematikaufgaben) und intrinsische Motivation nachgewiesen. Die Forschung zu den Effekten von Peer-Feedback (z.B. Gielen, Peeters, Dochy, Onghena, & Struyven, 2010) identifizierte in ähnlicher Weise die Wahrnehmung des Feedbackgebers als Autorität bzw. die Glaubwürdigkeit des Feedbacks als entscheidenden Faktor für die Feedbackeffektivität (Strijbos, Narciss, & Dünnebier, 2010). Die vorliegende Untersuchung greift diese beiden Ansätze auf, indem die wahrgenommene Feedbackeffektivität (d.h. wahrgenommene Nützlichkeit und Glaubwürdigkeit des Feedbacks) im Sinne von Hattie und Timperley (2007; siehe 1.1) als Mediator untersucht wird: Nehmen Lernende mit Fremdfedback die erhaltene Feedbackinformation als effektiver wahr im Vergleich zu den Lernenden, die sich ihre Feedbackinformationen selbst generieren? Und erklärt diese Divergenz in der wahrgenommenen Feedbackeffektivität die Vorteile externen gegenüber internem Feedback auf die Performanz?

1.4 Ziele der Studie

Die vorliegende Studie untersucht kompetenzielle Feedbackeffekte auf Performanz und wahrgenommene Feedbackeffektivität in Abhängigkeit von der Feedbackquelle (extern

versus intern). Angenommen wird, dass Fremdfeedback (extern) dem Selbstfeedback (intern) überlegen ist. Dabei steht die Forschungsfrage im Zentrum, worin diese Effekte auf die Performanz begründet sind.

Obwohl in beiden Feedbackbedingungen die Kriterien effektiven Feedbacks (siehe 1.1) berücksichtigt sind, wird angenommen, dass sich externes und internes Feedback bezüglich wahrgenommener Feedbackeffektivität unterscheiden. Diese Hypothese basiert auf den Befunden der Peer-Feedback-Forschung (vgl. Strijbos et al., 2010), wonach von einer Autorität gegebenes Feedback als effektiver wahrgenommen wird.

Vor allem aber wird in der vorliegenden Studie die wahrgenommene Feedbackeffektivität als potenzieller Mediator der Unterschiede zwischen kompetenziellem Fremd- und Selbstfeedback analysiert.

2 METHODE

Die experimentelle Studie untersuchte Effekte externen und internen kompetenziellen Feedbacks auf Performanz und wahrgenommene Feedbackeffektivität. Die Domäne war wissenschaftliches Denken im naturwissenschaftlichen Unterricht.

2.1 Stichprobe

Die Probanden waren Schülerinnen und Schüler aus vier 8. Klassen an Hamburger Schulen. Die Teilnahme erfolgte freiwillig. Die Analysestichprobe bestand aus $N = 50$ Schülerinnen und Schülern ($M = 14.58$ Jahre; $SD = 0.69$; 27 weiblich). Die Zuteilung der Probanden zu den zwei Feedbackbedingungen erfolgte randomisiert per Losverfahren.

2.2 Unabhängige Variable

Als Stufen der unabhängigen Variable *Feedbackquelle* fungierten eine Bedingung mit kompetenziellem Fremdfeedback und eine Bedingung mit kompetenziellem Selbstfeedback. Beide Feedbackbedingungen wurden in Form eines Rückmeldebogens im Papierformat umgesetzt. Dieser Rückmeldebogen enthielt eine Tabelle mit den schülergerecht formulierten fünf Kompetenzstufen wissenschaftlichen Denkens (Teilkompetenz „Experimente planen“), basierend auf dem Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung (Mayer, Grube, & Möller, 2008)²: „Bei der Planung eines Experiments hast du...: 1) ...gesagt, was du untersuchen willst. Du hast eine Variable genannt. Eine Variable ist die zu untersuchende Bedingung, die im Experiment verändert, variiert wird., 2) ...diese Variable verändert und gesagt, wie du die Veränderungen messen willst., 3) ...andere Einflussfaktoren, die du nicht untersuchen willst, gleich gehalten., 4) ...gesagt, an wie vielen Untersuchungsgegenständen, wie oft oder wie lange du das Experiment durchführen willst., 5) ...über mögliche Fehler oder Schwächen in deiner Planung nachgedacht und hast dann eine neue Untersuchung entworfen.“ Die dem Kompetenzmodell zugrundeliegende Niveaustufung ist nach zunehmender Komplexität graduiert und reicht von der „Untersuchung eines Faktors“, „Untersuchung von Zusammenhängen“, „Kontrollierte Untersuchung auf Basis von Konzeptverständnis“, „Elaborierte Untersuchung allgemeiner Zusammenhänge“ bis hin zum „Selbständigen Lösen von offenen Problemen“ (Mayer et al., 2008).

²Die Überlegungen des Kompetenzmodells naturwissenschaftlicher Erkenntnisgewinnung (Mayer et al., 2008) flossen in die Entwicklung des ESNaS-Modells („Evaluation der Standards in den naturwissenschaftlichen Fächern der Sekundarstufe 1“, vgl. Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010) ein. Am Ausgangsmodell von Mayer et al. (2008) wurde im Rahmen dieser Studie festgehalten, da sie Teil eines im Jahre 2008 vor der Veröffentlichung des ESNaS-Modells begonnenen Forschungsprojekts ist; um die Vergleichbarkeit der mehreren aufeinander aufbauenden Teilstudien des Forschungsprojekts zu gewährleisten, wurde immer dasselbe Kompetenzmodell verwendet.

In der *externen* kompetenziellen Feedbackbedingung wurde den Probanden durch Ankreuzen (5-Stufen-Skala) verdeutlicht, welche Kompetenzstufen sie erreicht hatten. Die Bewertungen wurden von der Versuchsleiterin durchgeführt. In der *internen* Feedbackbedingung schätzten die Probanden ihre Performanz selbst in Relation zu den dargebotenen fünf Kompetenzstufen ein. Die Probanden wurden in beiden Feedbackbedingungen darauf hingewiesen, dass die nicht angekreuzten Felder aufzeigen, wie die nächste Aufgabebearbeitung verbessert werden kann. Insgesamt ist gewährleistet, dass die Probanden beider Bedingungen identische Informationen (mit Ausnahme der Quelle der Einstufung) erhielten.

2.3 Abhängige Variablen

2.3.1 Performanz

Die erste abhängige Variable war die Performanz beim wissenschaftlichen Denken im Sinne der Güte der Bearbeitung von Aufgaben zur Planung von Experimenten. Die Performanz wurde mithilfe von zwei offenen Papier-Bleistift-Aufgaben (Mayer et al., 2008) erhoben. In den Aufgaben wurden wissenschaftliche Fragestellungen in alltagsnahen Situationen beschrieben, woraufhin die Lernenden jeweils ein Experiment planen sollten, welches die beschriebene Fragestellung beantwortet (z.B. „Tom möchte seiner Mutter zum Geburtstag Rosen schenken. In einer Gartenzeitschrift hat er gelesen, dass Rosen besser in Blumenerde wachsen als in Gartenerde. Bevor er seiner Mutter die Rosen schenkt, möchte er diese Aussage aus der Gartenzeitschrift prüfen. Plane ein Experiment, mit dem diese Aussage naturwissenschaftlich überprüft werden kann. Beschreibe das Experiment möglichst genau.“). Die offenen Antworten zu den Aufgaben wurden nach der Versuchsdurchführung erneut anhand des bestehenden Kodiermanuals von zwei hypothesen- und bedingungsblinden Ratern auf den Kompetenzstufen von 0 bis 5 kodiert. Die zufallskorrigierte Interrater-Reliabilität für Aufgabe 1 betrug $\kappa = 0.88$ und für die 2. Aufgabe $\kappa = 0.96$ (Cohen, 1960) und war somit

zufriedenstellend (Landis & Koch, 1977). In die statistische Hypothesenprüfung gingen die Rohwerte ein, auf die sich beide hypothesenblinden Rater geeinigt hatten. Die Übereinstimmung des kongruenten bedingungsblinden Ratings mit der Aufgabenbewertung, die von der Versuchsleiterin während der Versuchsdurchführung kodiert wurde, lag bei $\kappa = 0.89$ (Aufgabe 1). Die Aufgabenbewertung der Versuchsleiterin unterschied sich in zwei Fällen mit jeweils einem Punkt von dem bedingungsblinden Rating. Somit kann davon ausgegangen werden, dass das Fremdfeedback korrekte Feedbackinformationen beinhaltet. Korreliert man jedoch die Selbstbewertungen der Lernenden bei Aufgabe 1 mit dem zugehörigen bedingungsblinden Rating, so deutet Cohens Kappa eine niedrige Übereinstimmung an ($\kappa = 0.21$). Die weitere Analyse der niedrigen Korrektheit des Selbstfeedbacks mithilfe einer ANOVA zeigte, dass sich Lernende ($M = 3.56$; $SD = 1.29$) im Vergleich zu dem bedingungsblinden Rating ($M = 2.25$; $SD = 1.12$) signifikant überschätzten ($F(1,48) = 13.52$; $p < .01$; $\eta^2 = .16$). Inwiefern der Unterschied in der Genauigkeit der Feedbackinformationen einen Einfluss auf die abhängigen Variablen hatte, wird mithilfe der Kontrollvariable „Korrektheit des Feedbacks“ untersucht (siehe 2.4).

2.3.2 Wahrgenommene Feedbackeffektivität

Die wahrgenommene Feedbackeffektivität wurde nach der Feedbackintervention erfasst. Für die wahrgenommene Feedbackeffektivität wurden Items basierend auf der „Schülerperzipierten Lehrerbezugsnorm“ (SPLB, Schwarzer & Jerusalem, 1999) entwickelt. Der Fragebogen zur SPLB enthält Items, die nach erhaltenem Feedback die Wahrnehmung der Lernenden im Hinblick auf die Kriterien sozialer und individueller Bezugsnorm (vgl. Rheinberg, 2008) erfassen. Nach diesem Vorbild zielte die Skala „Wahrgenommene Feedbackeffektivität“ darauf ab, die Wahrnehmung der Lernenden im Hinblick auf die Kriterien effektiven Feedbacks nach Hattie und Timperley (2007, „Where am I going,

How am I going, Where to next“) zu erfassen: Die Items erhoben, inwiefern die Lernenden in dem erhaltenen Feedback wahrgenommen hatten, worin das Ziel der Aufgabe bestand („Die Rückmeldung zwischen den Aufgaben hat mir gezeigt, welche Antworten ich in der zweiten Aufgabe geben sollte.“), wie gut sie abgeschnitten hatten („Die Rückmeldung zwischen den Aufgaben hat mir gezeigt, wie ich bei der Planung von Experimenten abgeschnitten habe.“) und ob sie erkennen konnten, wie sie dem Ziel der Aufgabe näher kommen könnten („Die Rückmeldung zwischen den Aufgaben hat mir gezeigt, worauf ich bei der nächsten Aufgabe achten soll.“). Die Lernenden sollten auf einer Ratingskala von 1 = „stimmt nicht“ bis 4 = „stimmt genau“ angeben, wie sehr sie den drei Aussagen zustimmten.

2.4 Kontrollvariablen

Fünf Kontrollvariablen überprüften mögliche Unterschiede in den Ausgangsvoraussetzungen zwischen den Interventionsbedingungen. Erstens wurde als Indikator für die kognitiven Grundfähigkeiten die Figurenalogie-Skala des KFT 4-12+R (Heller & Perleth, 2000) mit den für die 8. Klasse normierten 25-Multiple-Choice-Items in einem Papier-Bleistift-Format eingesetzt. Zweitens erfassten drei Items das „Biologieinteresse“ der Lernenden (z.B. „Biologie interessiert mich.“), da die Aufgaben zur Planung von Experimenten in einen biologischen Kontext eingebettet waren. Drittens wurden zwei Items zur Erfassung der „Erfahrung mit der Planung von Experimenten“ (z.B. „Ich habe schon ähnliche Aufgaben zum Planen von Experimenten bearbeitet.“) eingesetzt. Die Items der Skalen „Biologieinteresse“ und „Erfahrung mit der Planung von Experimenten“ sollten von den Probanden auf einer Ratingskala von 1 = „stimmt nicht“ bis 4 = „stimmt genau“ eingeschätzt werden. Viertens wurden mögliche Performanzunterschiede vor der Feedbackintervention mithilfe des Pretests, d.h. der ersten Aufgabe zur Planung von Experimenten, analysiert. Fünftens wurde die „Korrektheit des Feedbacks“, d.h. der

Differenzwert zwischen der Fremd- und der Selbstbewertung in Relation zu Kompetenzstufen auf der einen Seite und der entsprechenden hypothesenblinden Raterkodierung auf der anderen Seite erfasst. Diese Kontrollvariable wurde eingesetzt, um die Alternativhypothese auszuschließen, dass unterschiedliche Effekte von Fremd- und Selbstfeedback ausschließlich auf eine möglicherweise korrektere Feedbackinformation im Fremdfeedback zurückzuführen sind (vgl. Kruger & Dunning, 2009). Durch die Subtraktion der Raterbewertungen von den Fremd- und Selbstbewertungen zeigt ein niedrigerer Wert eine korrektere Feedbackinformation an.

2.5 Versuchsablauf

Die Studie wurde in Gruppen mit maximal zehn Lernenden durchgeführt. Alle Instruktionen der Versuchsleiterin waren schriftlich fixiert und wurden vorgelesen. Die Lernenden erhielten zunächst einen Text als standardisierte Einführung in die Planung von Experimenten, die inhaltlich den Kompetenzstufen des Rückmeldebogens entsprach. Somit sollte gewährleistet werden, dass Probanden in beiden Versuchsbedingungen gleichermaßen mit dem Kompetenzmodell (Mayer et al., 2008) vertraut waren.

Es folgte die Planung des ersten Experiments durch die Lernenden. In der Bedingung *externes* Feedback kodierte die Versuchsleiterin die Antworten der offenen Aufgaben anhand eines bestehenden Kodiermanuals (Mayer et al., 2008), kreuzte im Rückmeldebogen die erreichte Kompetenzstufe an (siehe 2.2) und gab diesen an die Lernenden zurück. In der *internen* Feedbackbedingung füllten die Lernenden selbst den Rückmeldebogen aus. Anschließend planten die Probanden das zweite Experiment, schätzten die wahrgenommene Feedbackeffektivität ein und bearbeiteten die Figurenanalogie-Skala des KFT. Zum Schluss folgte ein Fragebogen, der demographische Items (Alter, Geschlecht) und die Items zu den Kontrollvariablen „Biologieinteresse“ und „Erfahrung mit der Planung von Experimenten“

beinhaltete. Zur Bearbeitung der Aufgaben standen jeweils maximal 10 Minuten zur Verfügung. Insgesamt betrug die Versuchsdurchführung ca. 60 Minuten.

3 ERGEBNISSE

3.1 Auswertungsmethode

Die Reliabilität der Skalen wurde als interne Konsistenz (Cronbachs Alpha) jeder Skala berechnet. Für die Analyse der Kontrollvariablen wurde eine multivariate Varianzanalyse gerechnet. Die Feedbackeffekte auf „Performanz“ und „Wahrgenommene Feedbackeffektivität“ wurden mit einfaktoriellen Varianzanalysen berechnet. Zur Untersuchung der dritten Forschungsfrage, inwiefern die wahrgenommene Feedbackeffektivität dem differenziellen Effekt der Feedbackquelle auf Performanz zugrundeliegt, wurden Regressionsanalysen gerechnet. Der Koeffizient der Feedbackquelle umfasste den Kontrast zwischen externem (1) und internem (-1) kompetenziellem Feedback. Der Performanzkoeffizient entsprach den Mittelwerten in der Posttestaufgabe (t2). Als Koeffizient für die Wahrnehmung der Effektivität des Feedbacks gingen die Mittelwerte der Skala „Wahrgenommene Feedbackeffektivität“ in die Analyse ein.

3.2 Reliabilität der Skalen

Die Reliabilität der eingesetzten Skalen war für die Performanz ($\alpha = .83$), die wahrgenommene Feedbackeffektivität ($\alpha = .82$), die kognitive Grundfähigkeit ($\alpha = .87$), das Biologieinteresse ($\alpha = .83$), und die Erfahrung mit der Planung von Experimenten ($\alpha = .82$) zufriedenstellend.

3.3 Kontrollvariablen

Für die Kontrollvariablen „Kognitive Grundfähigkeit“, „Biologieinteresse“, „Erfahrung mit der Planung von Experimenten“, Performanz in Aufgabe 1 und „Korrektheit des Feedbacks“ zeigte eine *MANOVA* keine signifikanten Unterschiede zwischen den Feedbackbedingungen (*Wilks Lambda* = .89; *Mult.F*(5,44) = 1.12; *p* = .36). Folglich kann hinsichtlich der Kontrollvariablen von gleichen Ausgangsbedingungen der beiden Experimentalgruppen ausgegangen werden.

3.4 Analyse der abhängigen Variablen

Im Folgenden sind die Analysen hinsichtlich der zwei abhängigen Variablen „Performanz“ und „Wahrgenommene Feedbackeffektivität“ dargestellt. Die Korrelation zwischen den beiden abhängigen Variablen betrug (*r* = .39; *p* = .01).

3.4.1 Performanz

Da im Pretest (Aufgabe 1) der Performanzunterschied zwischen den Feedbackbedingungen nicht signifikant war (siehe 3.3), wurde eine einfaktorielle *ANOVA* für den Posttest (Aufgabe 2) gerechnet. Diese weist einen signifikanten Unterschied zwischen den Interventionsbedingungen nach (*F*(1,48) = 6.05; *p* = .02; η^2 = .11): Lernende mit externem Feedback (*M* = 3.12; *SD* = 1.56) schnitten im Posttest signifikant besser ab als Lernende mit internem Feedback (*M* = 2.04; *SD* = 1.54). Damit kann Hypothese 1 als bestätigt gelten.

3.4.2. Wahrgenommene Feedbackeffektivität

Für die abhängige Variable „Wahrgenommene Feedbackeffektivität“ zeigte die einfaktorielle *ANOVA* ebenfalls einen signifikanten Unterschied zwischen den

Interventionsbedingungen ($F(1,48) = 5.61; p = .02; \eta^2 = .11$): Hypothese 2 entsprechend schätzten Lernende mit externem Feedback ihr Feedback signifikant als effektiver ein ($M = 3.28; SD = 0.77$) als Lernende mit internem Feedback ($M = 2.76; SD = 0.78$).

3.5 Mediationsanalyse

Erstens zeigte sich ein signifikanter Effekt von „Feedback“ auf „Performanz“ ($\beta = .34, t(48) = 2.46, p = .02$). Die experimentell variierte Feedbackquelle (Fremdfedback vs. Selbstfeedback) erklärte einen signifikanten Varianzanteil der abhängigen Variable Performanz ($R^2 = .11, F(1,48) = 6.05, p = .02$).

„Feedback“ war zweitens ein signifikanter Prädiktor für die Mediatorvariable „Wahrgenommene Feedbackeffektivität“ ($\beta = .32, t(48) = 2.37, p = .02$). Die experimentell variierte Feedbackvariable erklärte einen signifikanten Varianzanteil der Wahrnehmung der Effektivität des Feedbacks ($R^2 = .11, F(1,48) = 5.61, p = .02$).

Drittens konnte bezüglich des Zusammenhangs zwischen Mediatorvariable und abhängiger Variable „Wahrgenommene Feedbackeffektivität“ als signifikanter Prädiktor der „Performanz“ identifiziert werden ($\beta = .39, t(48) = 2.91; p = .01$). Die „Wahrgenommene Feedbackeffektivität“ erklärte einen signifikanten Varianzanteil der abhängigen Variable Performanz ($R^2 = .15, F(1,48) = 8.45, p = .01$).

In der vierten Regressionsanalyse mit beiden Prädiktoren „Feedback“ und „Wahrgenommene Feedbackeffektivität“ wurde der Effekt der unabhängigen Variable „Feedback“ auf „Performanz“ nicht signifikant ($\beta = .23, t(47) = 1.70; p = .10$), wohingegen der Mediatoreinfluss signifikant blieb ($\beta = .31, t(47) = 2.25; p = .03$).

Nach Baron und Kenny (1986) liegt somit eine totale Mediation vor: In allen durchgeführten Regressionsanalysen sagte der Prädiktor signifikant das Kriterium voraus, und der direkte Einfluss der unabhängigen Variable auf die abhängige Variable blieb nicht

signifikant, wenn die Mediatorvariable in die Regressionsanalyse aufgenommen wurde. Zur Überprüfung der Signifikanz des indirekten Effekts von „Feedback“ über „Wahrgenommene Feedbackeffektivität“ auf „Performanz“ wurde ein Sobel-Test (Sobel, 1982; MacKinnon, Warsi, & Dwyer, 1995) durchgeführt. Der Sobel-Test wurde signifikant ($t_{\text{Sobel}} = 1.65$; $p = .04$).

4 DISKUSSION

Die vorliegende Studie untersuchte Effekte kompetenziellen Feedbacks auf Performanz und wahrgenommene Feedbackeffektivität in Abhängigkeit von der Feedbackquelle, d.h. vergleichend zwischen Fremd- und Selbstfeedback.

Für die Performanz zeigte sich hypothesenkonform der schon nachgewiesene differenzielle Feedbackeffekt (Wollenschläger et al., 2012): Obwohl mit den Kompetenzstufen, auf denen die Performanz bewertet wurde, beiden Feedbackbedingungen dasselbe Kriterium der Leistungsbewertung zugrunde lag, schnitten Lernende mit kompetenziellem Fremdfeedback signifikant besser ab als Lernende, die ihre Performanz selbst in Relation zu Kompetenzstufen bewerteten.

Auch der erwartete Effekt der Feedbackquelle auf die wahrgenommene Feedbackeffektivität war feststellbar: Lernende mit Fremdfeedback nahmen die erhaltenen Feedbackinformationen signifikant effektiver wahr als Lernende, die sich selbst beurteilten.

Die Untersuchung des mediierenden Effekts der beiden Feedbackquellen auf die Performanz war die zentrale Forschungsfrage der vorliegenden Studie. Dabei konnte die wahrgenommene Feedbackeffektivität als Mediator identifiziert werden: das effektiver wahrgenommene Fremdfeedback führte zu dem signifikanten besseren Abschneiden der Lernenden. Dadurch wird deutlich, dass die *wahrgenommene* Feedbackeffektivität entscheidend ist für die *tatsächliche* Effektivität kompetenziellen Feedbacks. Somit konnte

der theoretisch angenommene Zusammenhang zwischen wahrgenommener Feedbackeffektivität und Performanz (z.B. Strijbos et al., 2010) empirisch gezeigt werden. Dieser Befund unterstreicht die Wichtigkeit der multidimensionalen Betrachtung von Feedback, dessen Effekte zwar auf Aufgabenebene wirksam und messbar werden, dessen Effektivität jedoch von internalen Prozessen des Feedbackempfängers vermittelt sind (Hattie & Wollenschläger, 2012).

Allerdings ist bezüglich des gefundenen totalen Mediationseffekts der wahrgenommenen Feedbackeffektivität zu betonen, dass das eingesetzte kompetenzielle Feedback sowohl in der Fremd- als auch in der Selbstbewertung die Kriterien effektiven Feedbacks umfasste. Die Schlussfolgerung der Mediation kann folglich nicht sein, dass eine Performanzsteigerung allein durch eine Illusion effektiven Feedbacks hervorgerufen werden kann. Die Schlussfolgerung sollte eher lauten, dass auch effektiv gestaltetes Feedback nur dann wirken kann, wenn es auch als effektiv wahrgenommen wird.

Bei der Ergebnisgeneralisierung sollten folgende Einschränkungen beachtet werden: Die Befunde entstanden in einem Setting, in dem das kompetenzielle Selbstfeedback, d.h. die Selbstbewertung in Relation zu Kompetenzstufen vorher nicht trainiert wurde. Selbstbewertung stellt für Lernende aber eine komplexe Herausforderung dar (vgl. Andrade & Du, 2007) und ist eine im Gegensatz zum Fremdfedback durch eine Lehrkraft weitaus weniger bekannte Situation (Adams & King, 1995). Möglicherweise schätzten die Lernenden in dieser für sie neuartigen Lernsituation ihre eigene Fähigkeit zur Selbstbewertung als gering ein (vgl. Hanrahan & Isaacs, 2001; Van Gennip, Segers, & Tillema, 2010), was in der Folge zu weniger effektiv wahrgenommenem Feedback führte. Dagegen scheinen Lernende den Lehrkräften eine hohe Bewertungskompetenz zuzusprechen (Adams & King, 1995; Strijbos et al., 2010) und folglich deren Feedback als effektiv zu erleben. Dieser mögliche Zusammenhang sollte in weiteren Studien spezifisch untersucht werden.

Obwohl die in der Studie gefundenen überschätzenden Selbstbewertungen für die negativen Performanzeffekte nicht entscheidend waren, wäre es in zukünftigen Studien lohnend, vor der Feedbackintervention die Genauigkeit wie auch vor allem die Effektivitätswahrnehmung des eigenen Feedbacks zu trainieren (vgl. Adams & King, 1995; Kasanen & Rätty, 2002). Durch ein solches Training hervorgerufene akkurate und effektiv wahrgenommene Selbstbewertungen in Relation zu den Kompetenzstufen könnten im Vergleich zu Fremdfedback möglicherweise noch aufgabenspezifischere Selbstreflexion und Metakognitionen (vgl. Cooper, 2006; Dochy, Segers, & Sluijsmans, 1999) anregen; unter diesen Voraussetzungen könnte die Hypothese dann sogar lauten, dass internes Feedback lernförderlicher sei als externes Feedback.

Des Weiteren ist zu beachten, dass in den bisher durchgeführten Studien der Einfluss der Lehrkraft konstant gehalten wurde. Für einen Transfer der Ergebnisse in den Schulalltag und die Erhöhung der ökologischen Validität sollten in einer Feldstudie die Lehrkraft und deren (tatsächliche und wahrgenommene) diagnostische Kompetenz (Südkamp, Kaiser, & Möller, in Druck) als Variablen einbezogen werden.

Zusammenfassend unterstreicht die Studie die Effektivität von kompetenziellem Fremdfedback auf die Performanz von Lernenden und betont die zentrale Rolle der Wahrnehmung der Feedbackeffektivität in der Feedbackverarbeitung.

5 LITERATUR

- Adams, C., & King, K. (1995): Towards a framework for student self-assessment. *Innovations in Education & Training International*, 32, 336-343. doi:10.1080/1355800950320405
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 159-181. doi:10.1080/02602930600801928
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238. doi:10.3102/00346543061002213
- Baron, K. E., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology*, 51, 1173-1182. doi:10.1037/0022-3514.51.6.1173
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, 18, 5-25. doi:10.1080/0969594X.2010.513678
- Bernholt, S., Parchmann, I., & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 217-243.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74. doi:10.1080/0969595980050102
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65, 245-281. doi:10.3102/00346543065003245
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. doi:10.1177/001316446002000104

- Cooper, D. (2006). *Talk about assessment: Strategies and tools to improve learning*. Toronto, ON: Thomson Nelson. Government of British Columbia.
- Dochy, F. , Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331-350. doi:10.1080/03075079912331379935
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20, 304-315. doi:10.1016/j.learninstruc.2009.08.007
- Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97, 21-34. doi:10.1080/00220670309596625
- Hanrahan, J., & Isaacs, G. (2001). Assessing self- and peer assessment: the students' views. *Higher Education Research and Development*, 20, 53-70. doi:10.1080/07294360123776
- Harks, B., Rakoczy, K., Hattie, J., Klieme E., & Besser, M. (2011). *Indirect feedback effects on mathematics achievement*. Manuskript eingereicht zur Publikation.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In P. Alexander, & R. E. Mayer (Eds.), *Handbook of research on learning and instruction* (pp. 249-271). New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi:10.3102/003465430298487

- Hattie, J., & Wollenschläger, M. (2012). *A conceptualization of feedback*. Manuskript eingereicht zur Publikation.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeits-Test für 4.-12. Klassen, Revision (KFT 4-12+R)*. Göttingen: Beltz.
- Jäger, R. S. (2008). Leistungsbeurteilung. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 324-336). Göttingen: Hogrefe.
- Kasanen, K., & Rätty, H. (2002). "You be sure now to be honest in your assessment": Teaching and learning self-assessment. *Social Psychology of Education, 5*, 313-328. doi:10.1023/A:1020993427849
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den naturwissenschaftlichen Fächern der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 135-153.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: Bundesministerium für Bildung und Forschung.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. *Zeitschrift für Pädagogik, 56*, 64-74.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284. doi:10.1037//0033-2909.119.2.254
- Köller, O., & Parchmann, I. (2012). Competencies: The German Notion of Learning Outcomes. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Making it tangible - Learning outcomes in science education* (pp. 165 - 185). Münster: Waxmann.

- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Journal of Psychology, 216*, 61-73. doi:10.1027/0044-3409.216.2.61
- Kruger, J., & Dunning, D. (2009). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134. doi:10.1037/0022-3514.77.6.1121
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211-232. doi:10.3102/00346543047002211
- Landis, G., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62. doi:10.1207/s15327906mbr3001_3
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms, & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 63-79). Innsbruck: Studienverlag.
- Möller, J., & Köller, O. (1996). Attributionen und Schulleistung. In J. Möller, & O. Köller (Hrsg.): *Emotionen, Kognitionen und Schulleistung* (S. 115-136). Weinheim: Beltz.
- Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

- Nolen, S. B. (1996). Why study? How reasons for learning influence strategy selection. *Educational Psychology Review*, 8, 335-355. doi:10.1007/BF01463938
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider, & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (S. 178–186). Göttingen: Hogrefe.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. doi:10.1007/BF00117714
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5, 77-85. doi:10.1080/0969595980050104
- Schecker, H., & Parchmann, I. (2006): Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45-66.
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, 18, 455-463. doi:10.1006/ceps.1993.1034
- Schwarzer, R., & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen*. Berlin: Freie Universität Berlin.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189. doi:10.3102/0034654307313795
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290-312). Washington: American Sociological Association.
- Stiggins, R., & Chapuis, J. (2006). What a difference a word makes: Assessment FOR learning rather than assessment OF learning helps students succeed. *Journal of Staff Development*, 27, 10-14.

- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291-303. doi:10.1016/j.learninstruc.2009.08.008
- Südkamp, A., Kaiser, J., & Möller, J. (in Druck). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*.
- Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learning and Instruction, 20*, 280-290. doi:10.1016/j.learninstruc.2009.08.010
- Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Zeitschrift für Pädagogik, 56*, 92-103.
- Wollenschläger, M., Hattie, J., Möller, J., & Harms, U. (2011a). *Competential feedback effects on performance: are they mediated by calibration?* Manuskript eingereicht zur Publikation.
- Wollenschläger, M., Möller, J., & Harms, U. (2011b). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift für Pädagogische Psychologie, 25*, 197-202. doi:10.1024/1010-0652/a000040
- Wollenschläger, M., Möller, J., & Harms, U. (2012). *External versus internal competential feedback: differential effects on performance and calibration?* Manuskript eingereicht zur Publikation.

DISKUSSION UND AUSBLICK

Zusammenfassend werden nun zunächst die Ergebnisse der vier Teilstudien sowie weitere Forschungsfragen unter Berücksichtigung bisheriger methodischer Einschränkungen dargestellt. Anschließend wird ein Gesamtfazit der empirischen Befunde gezogen und deren Relevanz und Implikationen für Forschung und Unterrichtspraxis skizziert.

1 DISKUSSION DER VIER TEILSTUDIEN

Die vier Teilstudien untersuchten allesamt Effekte kompetenziellen Feedbacks im schulischen Kontext und zwar im Bereich des naturwissenschaftlichen Unterrichts. Kompetenzielles Feedback entspricht der Leistungsbewertung und -rückmeldung in Relation zu den Stufen eines Kompetenzmodells.

In den vier Teilstudien wurde kompetenzielles Feedback mit einer Kontrollgruppe ohne Feedback (Studien 1 und 2) sowie mit anderen Feedbackarten wie sozialem Feedback (Studie 2) und internem Feedback (Studien 3 und 4) verglichen. Außerdem wurden Effekte kompetenziellen Feedbacks auf mehrere abhängige Variablen untersucht: zum einen auf die Performanz von Lernenden in der Domäne des wissenschaftlichen Denkens, bezogen auf den Teilbereich der Planung von Experimenten (Studien 1 bis 4). Darüber hinaus wurden Feedbackeffekte auf die Motivation (Studien 1 und 2), auf die Genauigkeit der Selbsteinschätzung der Lernenden (Studien 2 und 3) und auf die wahrgenommene Feedbackeffektivität (Studie 4) untersucht. Außerdem wurden Mediationsprozesse des Feedbackeffekts auf Performanz untersucht: als Mediatorvariablen wurden die Genauigkeit der Selbsteinschätzung (Studien 2 und 3) und wahrgenommene Feedbackeffektivität (Studie 4) in den Blick genommen.

Im Folgenden werden die Ergebnisse und das Fazit jeder Teilstudie dargestellt.

1.1 Studie 1

Die Forschungsfrage der ersten Studie lautete: Welche Effekte zeigt kompetenzielles Feedback im Vergleich zu keinem Feedback auf die Performanz und Motivation von Lernenden?

Die Ergebnisse zeigten, dass Lernende, die während der Bearbeitung von Aufgaben zum wissenschaftlichen Denken kompetenzielles Feedback erhielten, besser abschnitten als Lernende, die kein Feedback erhielten. Zudem zeigten sich tendenziell positive Effekte des kompetenziellen Feedbacks auf motivationale Variablen wie beispielsweise Interesse und Kompetenzerleben der Lernenden.

Diese Studie erbrachte erste Hinweise, dass Feedback in Relation zu den Stufen eines domänenspezifischen Kompetenzmodells positive Effekte auf den Lernprozess hat; diese Aussage gilt jedoch nur im Vergleich mit einer Kontrollgruppe ohne Feedbackinformation.

1.2 Studie 2

Da aufgrund von Studie 1 keine Aussage getroffen werden konnte, inwiefern kompetenzielles Feedback anderen Feedbackarten über- bzw. unterlegen ist, knüpfte Studie 2 genau an dieser Fragestellung an. Die Forschungsfragen lauteten erstens: Welche Effekte des kompetenziellen Feedbacks auf Performanz, Motivation (Kompetenzerleben) und Genauigkeit der Selbsteinschätzung zeigen sich im Vergleich zu einem sozialen Feedback und einer Kontrollgruppe ohne Feedbackinformation? Und zweitens: Vermittelt die Genauigkeit der Selbsteinschätzungen den angenommenen positiven Effekt kompetenziellen Feedbacks auf die Performanz von Lernenden?

Die Ergebnisse zeigten, dass Lernende in der kompetenziellen Feedbackbedingung signifikant besser abschnitten, sich selbst kompetenter wahrnahmen und ihre eigene Leistung signifikant genauer einschätzten als Lernende mit sozialem oder keinem Feedback. Die

Genauigkeit der Selbsteinschätzung konnte als partieller Mediator des Feedbackeffekts auf Performanz identifiziert werden: Kompetenzielles Feedback führte zu genaueren Selbsteinschätzungen der Lernenden; diese wiederum trugen zu einer Performanzsteigerung bei.

Folglich untermauerte die zweite Studie zum einen die Effektivität kompetenziellen Feedbacks und zeigte zum anderen, dass kompetenzielles Feedback sozialem Feedback (wie es häufig im Schulalltag eingesetzt wird) überlegen ist. Darüber hinaus zeigte sich drittens durch die Untersuchung des Mediationseffekts, dass die Genauigkeit der Schülerselbsteinschätzung eine zentrale Rolle im Hinblick auf die Effektivität von Feedback spielt.

1.3 Studie 3

Die dritte Studie umfasste eine systematische Analyse der Effekte eines als effektiv erwiesenen und für Fremd- und Selbstbewertung identischen Vergleichsmaßstabs: den Kompetenzstufen des kompetenziellen Feedbacks.

Die Forschungsfragen lauteten dabei erstens: Zeigen sich in der Selbstbewertung kompetenziellen Feedbacks die bereits für die kompetenzielle Fremdbewertung nachgewiesenen positiven Effekte auf Performanz und Genauigkeit der Selbsteinschätzungen im Vergleich zu keinem Feedback? Und zweitens: Ist der erwartete Unterschied in den Versuchsbedingungen durch die Genauigkeit der Selbsteinschätzungen begründet?

Lernende mit externem kompetenziellem Feedback (Fremdbewertung) schnitten signifikant besser ab und schätzten ihre eigene Leistung signifikant genauer ein als Lernende mit internem kompetenziellem Feedback (Selbstbewertung) oder keinem Feedback (Kontrollbedingung). Lernende mit internem kompetenziellem Feedback unterschieden sich in ihrer Performanz nicht von den Lernenden in der Kontrollbedingung. Die Genauigkeit der

Selbsteinschätzung konnte als partieller Mediator des Feedbackeffekts auf Performanz identifiziert werden: Ausschließlich kompetenzielles Fremdfeedback führte zu genaueren Selbsteinschätzungen der Lernenden; diese wiederum trugen zu einer Performanzsteigerung bei.

Somit identifizierte die dritte Studie erstens differenzielle Effekte kompetenzieller Fremd- und Selbstbewertung auf Performanz und Genauigkeit von Selbsteinschätzung und verdeutlichte zweitens die zentrale Wirkung von Selbsteinschätzungen in der Wirkung und Verarbeitung von Feedback.

1.4 Studie 4

Die vierte Studie ging den in Studie 3 identifizierten differenziellen Effekten kompetenziellen Fremd- und Selbstfeedbacks auf Performanz nach. Neben der Replikation dieser Befunde stand im Zentrum, eine mögliche Ursache der differenziellen Effekte zu untersuchen. Als mögliche Ursache wurde anknüpfend an Peer-Feedbackstudien die wahrgenommene Feedbackeffektivität in Betracht gezogen.

Dabei stellten sich folgende Forschungsfragen: (1) Nehmen Lernende mit Fremdfeedback (extern) die erhaltene Feedbackinformation als effektiver wahr im Vergleich zu Lernenden, die sich ihr Feedback selbst generieren (intern)? (2) Erklärt diese Divergenz in der wahrgenommenen Feedbackeffektivität die differenziellen Effekte externen versus internen kompetenziellen Feedbacks auf die Performanz?

Ergebnisse zeigten, dass Lernende mit kompetenziellem Fremdfeedback signifikant besser abschnitten und das Feedback signifikant effektiver wahrnahmen als Lernende in der internen kompetenziellen Feedbackbedingung. Die wahrgenommene Feedbackeffektivität medierte die Feedbackeffekte auf Performanz vollständig.

Die vierte Studie verdeutlichte die zentrale Rolle der wahrgenommenen Feedbackeffektivität in der Feedbackverarbeitung. Allerdings ist bezüglich des gefundenen totalen Mediationseffekts der wahrgenommenen Feedbackeffektivität zu betonen, dass das eingesetzte kompetenzielle Feedback sowohl in der Fremd- als auch in der Selbstbewertung die Kriterien effektiven Feedbacks umfasste. Die Schlussfolgerung der Mediation kann folglich nicht sein, dass eine Performanzsteigerung allein durch eine Illusion effektiven Feedbacks hervorgerufen werden kann. Die Schlussfolgerung der vierten Studie sollte eher lauten, dass effektiv gestaltetes Feedback nur dann wirken kann, wenn es auch als effektiv wahrgenommen wird. Dieses Fazit verdeutlicht weiteren Forschungsbedarf im Bereich der Feedbackforschung; auf diesen wird im folgenden Abschnitt detailliert und unter Berücksichtigung methodischer Einschränkungen der bisherigen Studien eingegangen, bevor die Befunde insgesamt auf ihre Relevanz und Implikationen hin diskutiert werden.

2 ANSCHLUSSFRAGESTELLUNGEN ZUKÜNFTIGER FORSCHUNG

Die vier Teilstudien der vorliegenden Dissertation sind in zweifacher Hinsicht als Basis zukünftiger Forschung denkbar: Potenzielle Folgestudien könnten erstens die Generalisierbarkeit der Befunde erhöhen und zweitens die Forschungsfragen auf andere Kontexte oder Forschungsbereiche transferieren.

2.1 Generalisierbarkeit der Befunde

Aufgrund des in den vier Teilstudien gewählten Untersuchungsdesigns gilt es, Einschränkungen hinsichtlich der Generalisierbarkeit der Befunde zu diskutieren und in weiteren Studien aufzugreifen.

In den vier Teilstudien wurden ein standardisiertes Untersuchungsdesign und eine Versuchsdurchführung im schulischen Setting, aber unter laborähnlichen Bedingungen

gewählt. Um Performanzeffekte ausschließlich auf die Feedbackintervention zurückführen zu können, wurde ein kurzer Interventionszeitraum analysiert und der Einfluss der Lehrkraft konstant gehalten. Für einen Transfer der Ergebnisse in den Schulalltag und die Erhöhung der ökologischen Validität sollten Effekte kompetenziellen Feedbacks in einer Längsschnittuntersuchung im Rahmen einer Feldstudie untersucht werden. Dabei sollte die Lehrkraft sowie deren diagnostische Kompetenz (Südkamp, Kaiser, & Möller, in Druck) als Variablen miteinbezogen werden.

Das Untersuchungsdesign umfasste in den zwei Teilstudien, die den Vergleich zwischen Fremd- und Selbstbewertung kompetenziellen Feedbacks betrafen (Studie 3 und 4), kein Training der Selbstbewertungsfertigkeiten von Lernenden. Selbstbewertung stellt für Lernende eine komplexe Herausforderung dar (vgl. Andrade & Du, 2007) und ist eine – im Gegensatz zum Fremdfeedback durch eine Lehrkraft – weitaus weniger bekannte Situation (Adams & King, 1995). Diese mögliche Konfundierung der Ergebnisse sollte in weiteren Studien spezifisch untersucht werden. Dabei wäre es lohnend, vor der Feedbackintervention die Genauigkeit der Schüler selbstbewertungen wie auch die Effektivitätswahrnehmung des eigenen Feedbacks zu trainieren (vgl. Adams & King, 1995; Kasanen & Rätty, 2002). Durch ein solches Training hervorgerufene akkurate und effektiv wahrgenommene Selbstbewertungen in Relation zu den Kompetenzstufen könnten im Vergleich zu Fremdfeedback möglicherweise noch aufgabenspezifischere Selbstreflexionen und Metakognitionen (vgl. Dochy, Segers, & Sluijsmans, 1999) anregen. Unter diesen Voraussetzungen würde die Hypothese dann lauten, dass internes kompetenzielles Feedback lernförderlicher ist als externes kompetenzielles Feedback. Auch diese Hypothese gilt es in weiteren Studien zu überprüfen.

Darüber hinaus ist zu erwähnen, dass die Ergebnisse der abhängigen Variablen Motivation und wahrgenommene Feedbackeffektivität mit zeitökonomischen und pro

Subskala (Interesse und Kompetenzerleben) nur drei Items umfassenden Skalen erfasst wurden. Dieses Vorgehen war für die vier Teilstudien der Dissertation gerechtfertigt, da es sich dabei um erste Untersuchungen kompetenziellen Feedbacks auf Motivation und wahrgenommene Feedbackeffektivität handelte. Die bisherigen Befunde könnten nun in Folgestudien durch den Einsatz von umfangreicheren (z.B. Deci & Ryan, 2000; Rheinberg, Vollmeyer, & Burns, 2001) oder multidimensionalen Skalen (z.B. Strijbos, Narciss, & Dünnebier, 2010) auf eine breitere Basis gestellt werden.

2.2 Transfer in andere Kontexte und Forschungsbereiche

Über die soeben beschriebene Auflösung von Konfundierungen oder die Beseitigung bisheriger methodischer Einschränkungen hinaus weisen die Ergebnisse der durchgeführten vier Teilstudien weitere Anschlussfragestellungen auf.

Das für die vier Teilstudien entwickelte und empirisch überprüfte kompetenzielle Feedback betraf die Kompetenz des wissenschaftlichen Denkens und enthielt Feedbackinformationen hinsichtlich der Güte von Aufgabenbearbeitungen, bezogen auf die Teilkompetenz „Experimente planen“ (Mayer, Grube, & Möller, 2008). Folglich beantwortete das bisher untersuchte kompetenzielle Feedback die Feedbackfragen *Where am I going?*, *How am I going?* und *Where to next?* (vgl. Hattie & Timperley, 2007) auf der Aufgabenebene. Weitere Studien könnten Effekte kompetenziellen Feedbacks auch auf einer anderen Feedbackebene untersuchen, beispielsweise bezogen auf die Selbstregulationsebene (Hattie & Timperley, 2007). Auf diesem *meta-task level* würden Feedbackinformationen auf die Selbstregulation der Lernenden wie auch deren Feedbackverarbeitungsstrategien abzielen (vgl. Hattie & Wollenschläger, 2012). Das in den vier Teilstudien gewählte Untersuchungsdesign könnte auch auf andere Teilkompetenzen wissenschaftlichen Denkens wie zum Beispiel „Daten analysieren/Schlussfolgerungen ziehen“ (Mayer (2007, S. 181)

transferiert werden. Dabei könnte das kompetenzielle Feedback aus aktuell entwickelten und überprüften naturwissenschaftlichen Kompetenzmodellen abgeleitet werden (vgl. Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010; Ramseier, Labudde, & Adamina, 2011). Zudem sollten in den Naturwissenschaften auch andere Kompetenzbereiche in den Blick genommen werden; hier würde sich zum Beispiel die Bewertungskompetenz mit ihrem dazugehörigen Kompetenzmodell anbieten (Eggert & Bögeholz, 2006).

Zum besseren Verständnis der der Feedbackeffektivität zugrundeliegenden Mechanismen könnten weitere mediierende oder moderierende Variablen des Feedbackeffekts auf Performanz untersucht werden. Zur Ableitung weiterer potenzieller Mediatoren und Moderatoren im Feedbackprozess bietet die *multidimensional conceptualization of feedback* (Hattie & Wollenschläger, 2012) Anknüpfungspunkte. Diese Konzeptualisierung verdeutlicht, dass sowohl Lernervoraussetzungen wie die Leistungsstärke und das Vorwissen der Lernenden (vgl. Labuhn, Zimmerman, & Hasselhorn, 2010) als auch deren selbstregulatorische Prozesse, z.B. die wahrgenommene Selbstwirksamkeitserwartung (vgl. Cleary & Zimmerman, 2004) die Wahrnehmung der Feedbackeffektivität und folglich auch die tatsächliche Feedbackeffektivität beeinflussen.

Im Hinblick auf eine potenzielle Folgestudie wäre eine Kombination kompetenziellen Feedbacks mit anderen instruktionalen Maßnahmen wie metakognitivem *prompting* (z.B. Hoffman & Spatariu, 2008) oder Lernen mit Beispielaufgaben (Salden, Alevan, Schwonke, & Renkl, 2010) denkbar, um mögliche additive oder interagierende Effekte der instruktionalen Maßnahmen einbeziehen zu können. Darüber hinaus könnten Effekte kompetenziellen Feedbacks, die bisher im schulischen Kontext überprüft wurden, auch in anderen Lehr-/Lernkontexten eingesetzt werden: beispielsweise in computergestützten Lernumgebungen (vgl. Narciss & Huth, 2004) oder in organisationalen Kontexten (vgl. Anseel, Van Yperen, Janssen, & Duyck, 2012).

Ein weiterer vielversprechender Ansatz zukünftiger Forschung wäre die Verknüpfung kompetenziellen Feedbacks mit der Peer-Feedbackforschung, und dies vor allem im Bereich des dort untersuchten Trainings zur Verbesserung der Peer-Feedbackqualität (Gan & Hattie, 2012). Kompetenzielles Feedback könnte in diesem Training als Instrument eingesetzt werden, um sowohl die Bewertungskriterien zwischen Feedbackgeber und -empfänger transparent zu machen als auch vor allem die Kriterien effektiven Feedbacks im Bewertungsprozess zwischen Peers implizit zu vermitteln.

3 GESAMTFAZIT

Die Untersuchung eines auf Stufen eines Kompetenzmodells basierenden Feedbacks beantwortet sowohl Forschungsfragen aus dem Bereich der Kompetenzforschung als auch aus der Feedbackforschung. So wurden in den vorliegenden vier Teilstudien erstmals Effekte eines auf einem Kompetenzmodell basierenden Feedbacks auf Individualebene (Schülerinnen und Schüler) standardisiert untersucht. Des Weiteren ist es gelungen, kompetenzielles Feedback, basierend auf theoretisch beschriebenen und in der Empirie identifizierten Kriterien effektiven Feedbacks, zu erstellen sowie dessen Effekte systematisch mit anderen Feedbackarten vergleichend im Schulkontext zu untersuchen. Die Kompetenzstufen eines Kompetenzmodells stellen im Falle des kompetenziellen Feedbacks einen externen, theoretisch fundierten und objektiven Standard dar. Diese Erweiterung der kriterialen Bezugsnorm wurde in den vier Teilstudien einer systematischen Untersuchung der Feedbackeffekte auf kognitive (Performanz), motivationale (Interesse und Kompetenzerleben) und metakognitive (Genauigkeit der Selbsteinschätzung und wahrgenommene Feedbackeffektivität) Variablen unterzogen. Durch den Vergleich zwischen kompetenzieller Fremd- und Selbstbewertung, welche beide auf denselben Kriterien (d.h. Stufen eines Kompetenzmodells) basieren, wurde zudem ein situationaler Faktor in Form der

Feedbackquelle (extern versus intern) als Einflussfaktor im Feedbackprozess untersucht. Diese multidimensionale Herangehensweise des Untersuchungsdesigns, sprich die Erhebung nicht nur kognitiver, sondern auch motivationaler, metakognitiver und situationaler Variablen ermöglichte eine systematische Analyse der den Feedbackeffekten zugrundeliegenden Mediationsprozesse.

Zusammenfassend entsprechen die vier Teilstudien der Dissertation auf Forschungsseite somit einer multidimensionalen, systematischen und kontrollierten Untersuchung eines aus Feedbacktheorien und empirischen Befunden neu abgeleiteten Feedbacktyps und dessen Effekten auf den Lernprozess; sie steuern dadurch Erkenntnisse im Bereich der Grundlagenforschung zu Feedbackeffekten bei. Diese Erkenntnisse können wiederum mit den in der Einleitung beschriebenen Modellen in Bezug gesetzt werden und so der weiteren theoretischen Modellentwicklung dienen. So wird deutlich, dass das Feedbackmodell von Hattie und Timperley (2007) um die wahrgenommene Feedbackeffektivität ergänzt werden sollte, wie dies daraufhin bei Hattie und Wollenschläger (2012) der Fall ist. Bezüglich des Feedbackmodells von Vollmeyer und Rheinberg (1998) bestärken die Befunde die angenommene Sicht der Motivation als abhängige Variable von Feedbackeffekten; diese Sicht wurde folglich in der *multidimensional conceptualization of feedback* (Hattie & Wollenschläger, 2012) einbezogen. Was das Modell von Butler und Winne (1995) betrifft, so wird deren Annahme externer und interner Feedbackquellen durch die empirischen Befunde gestützt; es wird jedoch deutlich, dass internes Feedback nicht nur auf Ziele und Strategien der Lernenden wirkt, sondern auch auf deren Feedbackwahrnehmung und Genauigkeit der Selbsteinschätzung.

Aufgrund der Anbindung des neu untersuchten Feedbacktyps an die Kompetenzforschung und den darin entwickelten Kompetenzmodellen wird die aktuell geforderte Outcome-Orientierung von Unterricht (vgl. Klieme et al., 2003) aufgegriffen und

weist auf die praktische Anschlussfähigkeit kompetenziellen Feedbacks hin. Auch inhaltlich ist kompetenzielles Feedback als empirisch nachgewiesenes effektives Feedback für den schulischen Kontext vielversprechend, da es einen für Lehrende wie auch Lernende anwendbaren, transparenten und objektiven Vergleichsstandard darstellt. So wird der Nutzen des hier untersuchten kompetenziellen Feedbacks im Bereich des Experimentalunterrichts oder auch in Schülerlaboren (Brandt, Möller, & Kohse-Höinghaus, 2008) offensichtlich.

Darüber hinaus verdeutlichte die in den Studien der Dissertation vorgenommene Analyse der der Feedbackeffektivität zugrundeliegenden Variablen (z.B. genaue Selbsteinschätzung und wahrgenommene Feedbackeffektivität), wie die Feedbackverarbeitung verbessert werden kann. Bezogen auf die Schulpraxis würde das bedeuten, dass die Lehrkraft diese feedbackverarbeitenden Variablen gezielt ansteuern kann. So sollten Lehrende darauf achten, dass Lernende im Vorfeld der Feedbackrezeption ihre eigene Performanz genau einschätzen lernen und die Feedbackinformationen, die sie erhalten werden, als wirksam betrachten.

Zusammenfassend wird deutlich, dass die im Bereich der Grundlagenforschung durchgeführten vier Teilstudien der Dissertation Erkenntnisse bereitstellen, aus welchen empirisch abgesicherte Empfehlungen für die Unterrichtspraxis abgeleitet werden können.

4 LITERATUR

- Adams, C., & King, K. (1995): Towards a framework for student self-assessment. *Innovations in Education & Training International*, 32, 336-343. doi:10.1080/1355800950320405
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 159-181. doi:10.1080/02602930600801928
- Anseel, F., Van Yperen, N. W., Janssen, O., & Duyck, W. (2011). Feedback type as a moderator of the relationship between achievement goals and feedback reactions. *Journal of Occupational and Organizational Psychology*, 84, 703-722. doi:10.1348/096317910X516372
- Brandt, A., Möller, J., & Kohse-Höinghaus, K. (2008). Was bewirken außerschulische Experimentierlabors? Ein Kontrollgruppenexperiment mit Follow up-Erhebung zu Effekten auf Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, 22, 5-12. doi:10.1024/1010-0652.22.1.5
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65, 245-281. doi:10.3102/00346543065003245
- Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A schoolbased program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, 41, 537-550. doi:10.1002/pits.10177
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268. doi:10.1207/S15327965PLI1104_01

- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331-350. doi:10.1080/03075079912331379935
- Eggert, S., & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz. Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 177-194.
- Gan, M., & Hattie, J. (2012). *The effects of explicit instructional support on formulating peer feedback at task, process, and self-regulation levels*. Manuskript eingereicht zur Publikation.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi:10.3102/003465430298487
- Hattie, J., & Wollenschläger, M. (2012). *A conceptualization of feedback*. Manuskript eingereicht zur Publikation.
- Hoffman, B., & Spatariu, A. (2008). The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology*, 33, 875-893. doi:10.1016/j.cedpsych.2007.07.002
- Kasanen, K., & Rätty, H. (2002). “You be sure now to be honest in your assessment”: Teaching and learning self-assessment. *Social Psychology of Education*, 5, 313-328. doi: 10.1023/A:1020993427849
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den naturwissenschaftlichen Fächern der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135-153.

- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: Bundesministerium für Bildung und Forschung.
- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173-194. doi:10.1007/s11409-010-9056-2
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger, & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung* (S. 177-186). Berlin: Springer.
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms, & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 63-79). Innsbruck: Studienverlag.
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Eds.), *Instructional design for multimedia learning* (pp. 181-195). Münster: Waxmann.
- Ramseier, E., Labudde, P., & Adamina, M. (2011). Validierung des Kompetenzmodells HarmoS Naturwissenschaften: Fazite und Defizite. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 7-33.
- Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. *Diagnostica*, 2, 57- 66.
- Salden, R., Alevin, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, 38, 289-307. doi:10.1007/s11251-009-9107-8

- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291-303. doi:10.1016/j.learninstruc.2009.08.008
- Südkamp, A., Kaiser, J., & Möller, J. (in Druck). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*.
- Vollmeyer, R., & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System. *Zeitschrift für Pädagogische Psychologie, 12*, 11-23.

LEBENS LAUF

NAME: Mareike Wollenschläger (verw. Schreiber, geb. Buck)

GEBURTS DATUM: 11.03.1981

GEBURTS SORT: Filderstadt

FAMILIENSTAND: verheiratet

NATIONALITÄT: deutsch

SCHULAU SBILDUNG:

1987-2000 Rudolf-Steiner-Schule Nürtingen

STUDIUM:

10/2001-06/2008 Psychologiestudium an der Eberhard-Karls-Universität
Tübingen

BERUFLICHE TÄTIGKEIT:

09/2008-06/2012 Wissenschaftliche Angestellte am Leibniz-Institut für die
Pädagogik der Naturwissenschaften und Mathematik (IPN)
an der Universität Kiel, Abteilung Didaktik der Biologie

09/2009-06/2012 Promovendin im BMBF-Nachwuchsförderprogramm
"Empirische Bildungsforschung"

seit 07/2012 Wissenschaftliche Angestellte am Leibniz-Institut für
Wissensmedien in Tübingen

FORSCHUNGSPRAKTIKUM:

11/2011-12/2011 Forschungsaufenthalt bei Prof. John Hattie im Melbourne
Education Research Institute, University of Melbourne

DRITTMITTELEINWERBUNG: BMBF (Nachwuchsförderung im Rahmenprogramm zur
Förderung der Empirischen Bildungsforschung):
„Effekte kompetenzieller Rückmeldung beim Experimentieren
im naturwissenschaftlichen Unterricht“
Laufzeit: 10/2009-06/2012

KIEL, IM JULI 2012