

The final publication is available at Hogrefe via DOI:

<http://dx.doi.org/10.1024/1010-0652.19.12.97>

Zeitschrift für Pädagogische Psychologie, 19, 97-106 © 2005 by  
Dickhäuser, O & Seidler, A. & Kölzer, M. This version of the  
article may not completely replicate the final version pub-  
lished in Zeitschrift für Pädagogische Psychologie. It is not  
the version of record and is therefore not suitable for cita-  
tion.

KOLUMNENTITEL: KEIN MENSCH KANN ALLES?

Kein Mensch kann alles?  
Effekte dimensionaler Vergleiche  
auf das Fähigkeitsselbstkonzept

Oliver Dickhäuser, Alexandra Seidler & Marika Kölzer  
Justus-Liebig-Universität Gießen

Autorenhinweis

Dr. Oliver Dickhäuser, Justus-Liebig-Universität Gießen, FB  
06, Otto-Behaghel-Straße 10F, 35394 Gießen. e-mail: oli-  
ver.dickhaeuser@psychol.uni-giessen.de

## Zusammenfassung

Referenzrahmenmodelle nehmen an, dass neben sozialen Vergleichen auch dimensionale Vergleiche Einfluss auf die Entstehung von Fähigkeitsselbstkonzepten nehmen. Haben Personen die Möglichkeit, die eigene Leistung in einem Bereich A mit schlechteren Leistungen in einem anderen Bereich B zu kontrastieren (dimensionaler Abwärtsvergleich), so soll dies zu einem Ansteigen des Fähigkeitsselbstkonzepts in Bezug auf A führen, ein dimensionaler Aufwärtsvergleich dagegen zu einem Absinken. In zwei Experimenten, in denen mit fingierter Leistungsrückmeldung gearbeitet wurde, zeigten sich tatsächlich Effekte dimensionaler Vergleichsinformation auf Fähigkeitsselbstkonzepte. Erstmals konnten im Experiment 2 sowohl positive Effekte dimensionaler Abwärtsvergleiche als auch negative Effekte dimensionaler Aufwärtsvergleiche nachgewiesen werden.

Schlagwörter: Leistung -- Selbstkonzept -- dimensionaler Vergleich

Nobody is perfect? Effects of dimensional comparisons on task-specific self-concepts

Frame of reference models assume that both, social and intra-individual (= "dimensional") comparisons affect the academic self-concept. Comparing one's achievement in a domain A with lower achievement in another domain B (dimensional downward comparison) is expected to lead to an increased academic self-concept concerning A, whereas upward comparisons will lead to decreases. The hypotheses were tested in two experiments in which faked performance feedback was used and task-specific self-concepts were assessed. There were effects of dimensional comparisons concerning the task-specific self-concept. Experiment 2 showed positive effects of dimensional downward comparisons as well as negative effects of dimensional upward comparisons.

Keywords: academic-achievement -- academic-self-concept -- dimensional-comparison

## 1 Theoretischer Hintergrund

Fähigkeitsselbstkonzepte wirken sich vielfältig auf das Erleben und Verhalten von Personen in Lern- und Leistungssituationen aus (Wigfield & Karpathian, 1991). Gesichertes Wissen über deren Determinanten eröffnet Ansatzpunkte, um Erleben und Verhalten von Personen in Lernkontexten mittelbar positiv zu beeinflussen. In der Literatur herrscht Einigkeit darüber, dass vorauslaufende Leistungen das Fähigkeitsselbstkonzept maßgeblich beeinflussen. Bezugsrahmentheorien nehmen dabei an, dass beobachtete eigene Leistungen zunächst in Beziehung zu einem Vergleichsmaßstab gesetzt werden müssen, um Aussagen über die Höhe eigener Fähigkeiten treffen zu können (Skaalvik & Skaalvik, 2002). Das von Marsh (1986, 1994) vorgeschlagene internal/external frame of reference Modell (I/E-Modell) geht davon aus, dass Personen sowohl externale (d.h. soziale) als auch internale (d.h. dimensionale) Vergleiche zur Bildung ihres Fähigkeitsselbstkonzepts verwenden. Beim sozialen Vergleich nutzt beispielsweise ein Schüler die Leistungen seiner Mitschüler, um die eigene Leistungsfähigkeit zu erschließen. Stellt er fest, dass die meisten seiner Mitschüler schlechter abschneiden als er selbst (sozialer Abwärtsvergleich), so sollte dies zu einer Erhöhung seines Fähigkeitsselbstkonzepts führen. Beim dimensionalen Vergleich nutzt ein Schüler seine Leistungen in einem Fach B, um seine Leistungsfähigkeit in A zu erschließen. Wird die Leistung im Fach A mit schlechteren Leistungen in einem anderen Bereich B kontrastiert (dimensionaler Abwärtsvergleich), so sollte dies zu einer Erhöhung des Fähigkeitsselbstkonzepts in Bezug auf A führen, ein Aufwärtsvergleich dagegen zu einem Absinken. Diese Effekte lassen sich insbesondere dann zeigen, wenn Personen davon ausgehen, dass

sich Fähigkeiten in den Bereichen A und B gegenseitig ausschließen (Möller, Pohlmann, Streblow & Kauffmann, 2002). Vor diesem Hintergrund wird verständlich, dass dimensionale Vergleiche "psychisch logisch" sein können, obwohl sie dazu führen, dass fachspezifische Selbstkonzepte weitaus deutlicher disparieren als die entsprechenden Leistungen.

Die meisten Arbeiten zur Bedeutung von Referenzrahmen bei der Selbstkonzeptgenese sind pfadanalytischer Natur (für einen Überblick siehe Möller & Köller, 2004). Es zeigen sich deutlich positive Regressionsgewichte bei der Vorhersage von Fähigkeitsselbstkonzepten in einem bestimmten Fach oder Bereich aufgrund von Leistungen in dem entsprechenden Fach oder Bereich. Diese werden als Hinweis auf soziale Vergleichsprozesse gewertet. Weiterhin zeigen sich mäßig negative Regressionsgewichte bei der Vorhersage von Fähigkeitsselbstkonzepten in einem bestimmten Bereich aufgrund von Leistungen in einem anderen Bereich oder Fach (üblicherweise werden der sprachliche und der mathematische Bereich herangezogen; zur Bedeutung unterschiedlicher Fächer vgl. Dickhäuser, 2003; Rost, Dickhäuser, Sparfeldt & Schilling, 2004). Diese werden als Hinweis auf dimensionale Vergleichsprozesse gewertet. Ob es aber tatsächlich die Vergleichsprozesse sind, die zu Veränderungen in den Fähigkeitsselbstkonzepten führen, bleibt hierbei unklar (vgl. Lüdtke, Köller, Artelt, Stanat & Baumert, 2002).

Ein anderes Vorgehen besteht darin, Schüler/-innen aufzufordern, eigene Leistungsergebnisse [etwa Noten von Klassenarbeiten] dimensional zu vergleichen und die Selbstkonzeptveränderung in Abhängigkeit des Vergleichs zu analysieren. Möller und Köller (1998) fanden bei Siebt- und Achtklässlern, dass die Residuen der fachspezifischen Selbstkonzepte der Schüler,

die sich dimensional aufwärts verglichen niedriger waren als die Residuen bei Schülern, die sich dimensional abwärts verglichen (Dickhäuser & Galfe, 2004, konnten allerdings bei Grundschulern entsprechende Befunde für dimensionale Vergleiche nicht erzielen). Die kausale Erklärungskraft dieser Studien bedarf aufgrund des quasi-experimentellen Designs der Ergänzung.

Studien, in denen dimensionale Vergleiche explizit manipuliert werden, ermöglichen hier einen stärkeren Test der Annahmen von Bezugsrahmentheorien, als pfadanalytische oder quasi-experimentelle Studien ihn leisten können (vgl. Lüdtke et al., 2002; Marsh, 1986). Es liegen jedoch kaum entsprechende Studien vor. Was den Einfluss sozialer Vergleiche betrifft, so konnte Meyer (zsf. 1984) in mehreren Experimenten zeigen, dass Fähigkeitsselbstkonzepte erfolgreich durch sozial vergleichende Rückmeldung beeinflusst werden können. Erstmals haben Möller (2000) sowie Möller und Köller (2001a) Effekte dimensionaler Vergleiche auf aufgabenspezifische Fähigkeitsselbstkonzepte experimentell untersucht. Die Versuchspersonen bearbeiteten zwei unterschiedliche Aufgaben A und B und erhielten manipulierte dimensional vergleichende Rückmeldung. Es zeigte sich in Studie 2 von Möller und Köller (2001a), dass dimensionale Aufwärtsvergleiche zu einem Absinken des Fähigkeitsselbstkonzepts führen. Der erwartete positive Effekt dimensionaler Abwärtsvergleiche blieb jedoch aus. Letzteres ist vor dem Hintergrund des I/E-Modells (Marsh, 1986) unerwartet, demzufolge dimensionale Abwärtsvergleiche zu einem Anstieg des Fähigkeitsselbstkonzepts in der Referenzaufgabe führen sollten, Aufwärtsvergleiche dagegen zu einem Absinken.

Die vorliegende Arbeit will diese wenigen Arbeiten zum Einfluss dimensionaler Vergleiche auf das Fähigkeitsselbstkonzept ergänzen. Dabei sollen Annahmen von Bezugsrahmentheorien einer Prüfung unterzogen und offene Fragen geklärt werden. Studie 1 der vorliegenden Arbeit geht daher der Frage nach, ob sich der Befund der nicht gegengleichenden Effekte von Aufwärts- vs. Abwärtsvergleichen replizieren lässt. Es wird erwartet, dass dimensionale Vergleiche folgende Effekte auf das Fähigkeitsselbstkonzept haben: Nach dimensional aufwärtsvergleichen sinkt das Fähigkeitsselbstkonzept ab (Replikation der Befunde von Möller & Köller, 2001a), nach Abwärtsvergleichen dagegen steigt es an (dieser Effekt wird aufgrund des Modells von Marsh erwartet, konnte jedoch bislang experimentell nicht gezeigt werden).

## 2 Studie 1

### 2.1 Methode

Studie 1 stellt in Teilen eine Replikation der Arbeit von Möller (2000) dar, daher wurde ein ähnliches Design verwendet.

Stichprobe. Die Stichprobe bestand aus 61 Studierenden (davon 35 weiblich) verschiedener Fachbereiche der Justus-Liebig-Universität Gießen und der Fachhochschule Gießen/Friedberg. Das Alter lag zwischen 21 und 26 Jahren ( $M = 23.0$ ). Die Versuchspersonen (Vpn) wurden durch Aushänge und persönliche Ansprache angeworben. Die Teilnahme (Dauer etwa 60 Minuten) wurde nicht vergütet, lediglich Psychologiestudierende erhielten einen Teilnahmenachweis.

Überblick und Versuchsplan. Zu Beginn erhielten die Vpn Beispielaufgaben zweier unterschiedlicher Typen A und B. In Bezug auf diese Aufgaben wurden die Fähigkeitsselbstkonzepte erfasst. Anschließend erfolgte die Aufgabenbearbeitung. Nach

der Bearbeitung erhielten die Vpn nach Zufall eine dimensional abwärts oder aufwärts vergleichende Ergebnisrückmeldung. Bei der am darauffolgenden Tag stattfindenden erneuten Befragung wurden wiederum die Fähigkeitsselbstkonzepte für die Aufgaben erfasst. Der Untersuchung lag somit ein 2 [Messzeitpunkt] x 2 [Ergebnisrückmeldung: aufwärts vs. abwärts] faktorieller Versuchsplan zugrunde<sup>1</sup>. Abhängige Variablen waren die Fähigkeitsselbstkonzepte für A und B. Zusätzlich wurde die Leistung in den Aufgaben erfasst.

Material und Durchführung. Den Vpn wurde zum ersten Messzeitpunkt mitgeteilt, sie nähmen an einer Untersuchung teil, in welcher ein neuer Leistungstest für den deutschen Sprachraum normiert würde. Gegenstand der aktuellen Normierung seien aber nur die Aufgaben A („Figurenalogien“), während die Aufgaben B („Wortanalogien“) lediglich zu Vergleichszwecken mit untersucht würden. Dieser Hinweis sollte sicherstellen, dass die Vpn die nachfolgende Rückmeldung tatsächlich aus der Sicht von Aufgabe A als Aufwärts- bzw. Abwärtsvergleich interpretieren.

Die Vpn machten demografische Angaben und notierten ein Codewort auf ihrem Fragebogen. Dann wurden jeweils vier Beispielaufgaben des Bereichs Figurenalogien aus dem Kognitiven Fähigkeitstest KFT 4-13 (Heller, Gädicke & Weinläder, 1985), Form A, Aufgaben N2, Nr. 16, 40, 46 und 47 sowie des Bereichs Wortanalogien (Form B, Aufgaben V4, Nr. 4, 15, 21 und 43) präsentiert und erläutert. Die gewählten Aufgaben entsprechen denen der Studie von Möller (2000).

Danach wurden die aufgabenspezifischen Fähigkeitsselbstkonzepte für beide Aufgaben erfasst. Hierzu wurden jeweils fünf Items verwendet, welche das Fähigkeitsselbstkonzept abso-

lut, d.h. ohne Thematisierung von Bezugsnormen des Vergleichs erfassen (vgl. Dickhäuser, Schöne, Spinath & Stiensmeier-Pelster, 2002). Die Items lauteten für Figurenanalogien [bzw. Wortanalogien]: „Ich bin für die Aufgaben Figurenanalogien [Wortanalogien]...“ „nicht begabt“ [1] vs. „sehr begabt“ [5], „Die Aufgaben Figurenanalogien [Wortanalogien] neu zu lernen fällt mir...“ „schwer“ [1] vs. „leicht“ [5], „Ich bin für die Aufgaben Figurenanalogien [Wortanalogien]...“ „nicht intelligent“ [1] vs. „sehr intelligent“ [5], „Ich kann bei den Aufgaben Figurenanalogien [Wortanalogien]...“ „wenig“ [1] vs. „viel“ [5] sowie „Die Aufgaben Figurenanalogien [Wortanalogien] fallen mir...“ „schwer“ [1] vs. „leicht“ [5]. Die internen Konsistenzen in der vorliegenden Stichprobe waren sehr gut (Figurenanalogien t1 und t2: .95, Wortanalogien t1: .95, t2: .96).

Dann wurden wie bei Möller (2000) folgende Aufgaben zur Bearbeitung vorgegeben: „Figurenanalogien“ Items Nr. 51-55 sowie 61-70 (Form A) des Untertests N2, „Wortanalogien“ Items Nr. 21-24 und 33-50 (Form A) sowie Nr. 45-48 (Form B) des Untertests V4 aus dem KFT (Heller et al., 1985). Für jede der Aufgabengruppen standen 5 Minuten Bearbeitungszeit zur Verfügung. Die unterschiedliche Menge an Aufgaben liegt darin begründet, dass die Lösung eines Items der Figurenaufgaben mehr Zeit in Anspruch nimmt als die Lösung eines Items der Wortaufgaben. Die unterschiedliche Aufgabenmenge erweckt so bei den Versuchspersonen den Eindruck, sowohl in den Wort- als auch in den Figurenaufgaben relativ zur gegebenen Aufgabenmenge gleich viele Aufgaben lösen zu können.

Im Anschluss hieran wurden die Aufgabenblätter von einer Konföderierten eingesammelt und angeblich ausgewertet. Die Vpn bearbeiteten derweil hypothesenirrelevante Fragebögen und be-

kamen nach etwa 15 Minuten die Ergebnisse der angeblichen Testauswertung schriftlich auf einem mit dem Codewort gekennzeichneten Rückmeldebogen mitgeteilt. Die Ergebnisrückmeldung lautete wie folgt: „Du hast bei Aufgabe A (Figurenanalogien) im Vergleich zu Aufgabe B (Wortanalogien) wesentlich schlechter / schlechter / gleich / besser / wesentlich besser abgeschnitten.“ In der Bedingung „Abwärtsvergleich“ war die Antwortalternative „besser“ angestrichen, in der Bedingung „Aufwärtsvergleich“ die Antwortalternative „schlechter“.

Der zweite Messzeitpunkt erfolgte einen Tag später; die Studie wurde von einer anderen Versuchsleiterin durchgeführt und wurde den Vpn als Studie zum Verhalten in Wettbewerbssituationen vorgestellt. Den Vpn wurde mitgeteilt, dass aus Ökonomiegründen für diese Studie die gleichen Aufgaben verwendet würden wie in der Untersuchung am Vortag. Die Vpn sollten (als Manipulationscheck) zunächst angeben wie sie am Vortag in der Aufgabe A relativ zu Aufgabe B abgeschnitten haben. Dazu kreuzten sie eine der folgenden Alternativen an: „Ich habe bei Aufgabe A (Figurenanalogien) im Vergleich zu Aufgabe B (Wortanalogien) wesentlich schlechter / schlechter / gleich / besser / wesentlich besser abgeschnitten.“

Dann wurde den Vpn angekündigt, dass sie gleich eine bestimmte Menge an Aufgaben zur Bearbeitung bekommen würden. Anschließend wurden erneut die aufgabenspezifischen Fähigkeits-selbstkonzepte für die Aufgaben Figurenanalogien und Wortanalogien anhand der gleichen Items wie zum ersten Messzeitpunkt erfasst. Schließlich wurden die Versuchspersonen aufgeklärt.<sup>2</sup>

## 2.2 Ergebnisse

Um zu überprüfen, ob die Manipulation der Leistungsrückmeldung erfolgreich war, wurden die Antworten in Bezug auf den

Manipulationscheck je nach Rückmeldung ausgezählt. Eine Versuchsperson beantwortete den Manipulationscheck nicht und wurde aus der weiteren Auswertung ausgeschlossen. Die Manipulation war in allen übrigen Fällen erfolgreich: In der Bedingung „Abwärtsvergleich“ gaben alle Versuchspersonen die Antwortalternative „besser“ an, in der Bedingung „Aufwärtsvergleich“ wählten alle Versuchspersonen die Antwortalternative „schlechter“.

Um zu überprüfen, ob sich die Leistung in den Aufgaben Figurenalogien und Wortanalogien für die beiden Experimentalgruppen unterschied, wurde eine einfaktorielle multivariate Varianzanalyse mit dem Faktor Ergebnismeldung [aufwärts/abwärts] gerechnet. Abhängige Variablen war die Leistung in den Aufgaben Figurenalogien und Wortanalogien (jeweils Anzahl der richtig gelösten Aufgaben). Es zeigte sich kein Effekt der Rückmeldung ( $F < 1.25$ ).

Die Stabilität des Fähigkeitsselbstkonzepts für den Test Figurenalogien (Korrelation  $t_1, t_2$ ) betrug  $r = .62$ , für das Fähigkeitsselbstkonzept Wortanalogien ergab sich eine Stabilität von  $r = .64$ . Die Stabilitäten unterschieden sich nicht nach Rückmeldebedingung (alle  $z$ -Werte  $< 1.17$ ).

Um zu überprüfen, ob die Veränderungen im Fähigkeitsselbstkonzept für den Test Figurenalogien von  $t_1$  zu  $t_2$  je nach Rückmeldung unterschiedlich sind, wurde eine Varianzanalyse mit den Faktoren Ergebnismeldung und (als Messwiederholungsfaktor) Testzeitpunkt gerechnet. Abhängige Variable war das Fähigkeitsselbstkonzept für den Test Figurenalogien. Es zeigten sich keine Haupteffekte ( $F$ -Werte  $< 1.48$ ), jedoch ergab sich wie erwartet eine signifikante Interaktion von Messwiederholungsfaktor und Rückmeldung,  $F(1; 58) = 5.55$ ,  $p < .05$ ,

$\eta^2 = .09$ . Die entsprechenden Mittelwerte sind separat für die beiden Rückmeldebedingungen und Messzeitpunkte in Tabelle 1 dargestellt. Wie separate t-Tests zeigten, fiel von t1 zu t2 das Fähigkeitsselbstkonzept beim Aufwärtsvergleich signifikant ab,  $t(31) = 2.79$ ,  $p < .01$ . Der Anstieg von t1 zu t2 war dagegen in der Bedingung Abwärtsvergleich statistisch nicht signifikant,  $t < 0.97$ .

Schließlich wurde eine weitere Varianzanalyse mit Messwiederholung gerechnet. Faktoren waren wiederum Ergebnismeldung sowie der Testzeitpunkt als Messwiederholungsfaktor. Abhängige Variable war das Fähigkeitsselbstkonzept für den Test Wortanalogien. In dieser Analyse zeigte sich lediglich ein Haupteffekt des Faktors Rückmeldung auf das Fähigkeitsselbstkonzept Wortanalogien,  $F(1; 58) = 4.87$ ,  $p < .05$ ,  $\eta^2 = .08$ . Dieser beruhte darauf, dass das Fähigkeitsselbstkonzept zu beiden Messzeitpunkten in der Bedingung Aufwärtsvergleich höher war als in der Bedingung Abwärtsvergleich (vgl. Tabelle 1). Alle übrigen Effekte waren nicht signifikant (alle  $F$ -Werte  $< 0.48$ ).

### 2.3 Diskussion

Studie 1 konnte wie erwartet nachweisen, dass dimensionale Vergleichsinformation Veränderungen im Fähigkeitsselbstkonzept bewirkt. Es zeigte sich, dass es nach Aufwärtsvergleichen zu einem Absinken des Fähigkeitsselbstkonzepts kommt, während nach Abwärtsvergleichen kein analoger Anstieg im Fähigkeitsselbstkonzept zu verzeichnen war. Das Ausbleiben des Effekts für Abwärtsvergleiche repliziert zwar die Ergebnisse von Möller und Köller (2001a, Studie 2), der Nullbefund war jedoch vor dem Hintergrund der zugrunde liegenden Bezugsrahmentheorien nicht zu erwarten. Als Erklärung für das Ausbleiben des

Effekts führen Möller und Köller (2001a) an, dimensionale Abwärtsvergleiche blieben ohne Auswirkungen, wenn sie in Situationen vorgenommen werden, die kaum bedrohlich für das Selbst einer Person sind (wie dies möglicherweise in einem experimentellen Setting der Fall ist, nicht aber in realen Leistungs- und Rückmeldesituationen wie etwa der Schule). Möglicherweise trifft jedoch eine andere, einfachere Erklärung zu: In der vorliegenden Studie 1 erhielten die Versuchspersonen ebenso wie bei Möller und Köller (2001a) zunächst Beispielaufgaben, aufgrund derer das Selbstkonzept erfasst wurde. Dann erfolgte die Bearbeitung der Aufgaben und die manipulierte Rückmeldung. Das Aufgabenmaterial enthielt einfache Beispielaufgaben, während die zu bearbeitenden Aufgaben im Vergleich dazu deutlich schwerer waren<sup>3</sup>. Diese unterschiedliche Aufgabenschwierigkeit kann möglicherweise das Ausbleiben des Effekts dimensionaler Abwärtsvergleiche erklären. Liegen nämlich der Fähigkeitseinschätzung zum ersten Testzeitpunkt leichtere Aufgaben zugrunde als zum zweiten, sollten alle Personen eine generelle Tendenz zum Absenken ihrer Selbsteinschätzungen zeigen, welche dann die Effekte der dimensional vergleichenden Rückmeldung überlagert. Diese Vermutung soll in Studie 2 überprüft werden.

### 3 Studie 2

#### 3.1 Methode

Stichprobe. Die Stichprobe bestand aus 60 Studierenden (davon 13 männlich) der Justus-Liebig-Universität Gießen im Alter zwischen 19 und 38 Jahren ( $M = 21.7$ ). Die Vpn wurden während regulärer Lehrveranstaltung zur Teilnahme angeworben. Die Teilnahme dauerte etwa 60 Minuten und wurde mit € 5 vergütet, Psychologiestudierende erhielten statt des Geldes einen Teilnahmenachweis.

Überblick und Versuchsplan. Das Design der Studie 2 war mit dem der Studie 1 vergleichbar (Abwandlungen s.u.). Es handelte sich um einen 2 [Messzeitpunkt] x 2 [Ergebnisrückmeldung: aufwärts vs. abwärts] faktoriellen Versuchsplan<sup>4</sup>. Zusätzlich zu den in Studie 1 erfassten abhängigen Variablen wurde als Kontrollvariable das allgemeine studienbezogene Fähigkeitsselbstkonzept erfasst.

Material und Durchführung. Den Vpn wurde wie in Studie 1 mitgeteilt, sie nähmen an einer Untersuchung teil, in welcher ein neuer Leistungstest für den deutschen Sprachraum entwickelt würde. Der Begriff "Normierung" wurde jedoch vermieden, um nicht den Fokus auf soziale Vergleiche zu lenken.

Als Beispielaufgaben dienten nunmehr drei Aufgaben des Bereichs Figurenanalogien (KFT, Aufgaben N2, Items B51, B61, B66) sowie drei Aufgaben des Bereichs Wortanalogien (KFT, Aufgaben V4, Items B21, B33, A45). Diese Aufgaben entsprechen hinsichtlich ihrer Itemschwierigkeit gemäß der Testkonstruktion des KFT den später zu bearbeitenden Aufgaben.

Nach der Darstellung der jeweiligen Beispielaufgaben erfolgte die Erfassung der aufgabenspezifischen Fähigkeits-selbstkonzepte anhand der gleichen Items wie in Studie 1, allerdings wurde zum Zweck einer besseren Differenzierung eine Antwortskala mit einem Range von 1 bis 7 verwendet (die internen Konsistenzen in der vorliegenden Stichprobe waren sehr gut, Figurenanalogien t1: .96, t2: .95, Wortanalogien t1: .94, t2: .95). Zusätzlich wurde im Anschluss als Kontrollvariable das allgemeine studienbezogene Fähigkeitsselbstkonzept anhand von fünf Items der Skala von Dickhäuser et al. (2002) erfasst. Ein Beispiel lautete „Meiner Meinung nach bin ich...“ „nicht in-

telligent“ [1] vs. „sehr intelligent“ [7]. Die interne Konsistenz der Skala betrug  $\alpha = .87$ .

Die anschließend zu bearbeitenden Aufgaben, die dabei zur Verfügung stehende Zeit sowie das Vorgehen bei der Ergebnisrückmeldung waren identisch zu Studie 1. Jedoch erfolgte der zweite Befragungszeitpunkt gleich im Anschluss. Auch hier wurde die zweite Befragung den Vpn als Studie zum Verhalten in Wettbewerbssituationen vorgestellt. Wie zu t1 wurden dann die aufgabenspezifischen Fähigkeitsselbstkonzepte erfasst. Außerdem gaben die Vpn wie in Studie 1 - diesmal jedoch erst nach der Erfassung der Selbstkonzepte - an, wie sie im ersten Teil des Versuchs in der Aufgabe A relativ zu Aufgabe B abgeschnitten hatten.

### 3.2 Ergebnisse

Die Manipulation der Leistungsrückmeldung war in allen Fällen erfolgreich: In der Bedingung „Abwärtsvergleich“ gaben alle Versuchspersonen die Antwortalternative „besser“ an, in der Bedingung „Aufwärtsvergleich“ wählten alle Versuchspersonen die Antwortalternative „schlechter“.

Mittels MANOVA wurde geprüft, ob sich die Leistung in den Aufgaben für die beiden Experimentalgruppen unterschied. Faktor war die Rückmeldung, abhängige Variablen waren die Anzahl der richtig gelösten Aufgaben in den Figurenanalogien sowie in den Wortanalogien. Es zeigten sich kein Effekt der Rückmeldung ( $F < 0.05$ ). Um zu überprüfen, ob sich das allgemeine Fähigkeitsselbstkonzept für die Experimentalgruppen unterschied, wurde eine einfaktorielle Varianzanalyse mit dem Faktor Rückmeldung und dem allgemeinen Fähigkeitsselbstkonzept als abhängige Variable gerechnet. Es zeigte sich kein Effekt der Rückmeldung ( $F < 1.28$ ).

Die Stabilität des Fähigkeitsselbstkonzepts für den Test Figurenalogien lag bei  $\underline{r} = .74$ , für das Fähigkeitsselbstkonzept Wortanalogien ergab sich eine Stabilität von  $\underline{r} = .61$ . Auch hier unterschieden sich die Stabilitäten nicht nach Rückmeldebedingung (alle z-Werte  $< 0.66$ ).

Um zu überprüfen, ob die Veränderungen im Fähigkeitsselbstkonzept für den Test Figurenalogien von t1 zu t2 je nach Rückmeldung unterschiedlich ausfielen, wurde eine Varianzanalyse mit Messwiederholung gerechnet. Faktoren waren Rückmeldung sowie Testzeitpunkt als Messwiederholungsfaktor. Abhängige Variable war das Fähigkeitsselbstkonzept für den Test Figurenalogien. Es zeigten sich keine Haupteffekte ( $\underline{F} < 0.91$ ), jedoch ergab sich wie erwartet eine signifikante Interaktion von Messwiederholungsfaktor und Rückmeldung,  $\underline{F} (1; 58) = 16.33$ ,  $\underline{p} < .001$ ,  $\eta^2 = .22$ . Wie separate t-Tests zeigen, fiel das Fähigkeitsselbstkonzept nach einem dimensionalen Aufwärtsvergleich signifikant ab,  $\underline{t} (29) = 2.08$ ,  $\underline{p} < .05$ , während es nach einem dimensionalen Abwärtsvergleich statistisch signifikant anstieg,  $\underline{t} (29) = -3.72$ ,  $\underline{p} < .01$ . Die Mittelwerte getrennt nach Testzeitpunkt und Rückmeldung sind in Tabelle 2 dargestellt.

Schließlich wurde eine weitere Varianzanalyse gerechnet. Faktoren waren Ergebnissrückmeldung und Testzeitpunkt (als Messwiederholungsfaktor). Abhängige Variable war das Fähigkeitsselbstkonzept für den Test Wortanalogien. In dieser Analyse zeigte sich lediglich ein Haupteffekt des Messwiederholungsfaktors auf das Fähigkeitsselbstkonzept Wortanalogien,  $\underline{F} (1; 58) = 28.73$ ,  $\underline{p} < .001$ ,  $\eta^2 = .33$ . Dieser beruhte darauf, dass die Fähigkeitsselbstkonzepte in beiden Bedingungen zum ersten Messzeitpunkt höher waren, als zum zweiten Messzeit-

punkt (vgl. Tabelle 2). Alle übrigen Effekte waren nicht signifikant (alle  $F$ -Werte  $< 2.78$ ).

#### 4 Zusammenfassende Diskussion

In der vorliegenden Arbeit wurden Annahmen aus der Literatur zum Einfluss von dimensional Vergleichen auf Fähigkeitsselbstkonzepte getestet. Mit ihrem experimentellen Design bietet unsere Studie eine wertvolle Ergänzung zu den bislang vorwiegend pfadanalytischen Arbeiten zur Bedeutung dimensionaler Vergleiche (vgl. Lüdtke et al., 2002). Im Experiment können die Vergleichsprozesse – anders als in pfadanalytischen Designs – manipuliert und ihr Einfluss auf Veränderungen im Fähigkeitsselbstkonzept getestet werden. Das Problem pfadanalytischer Testungen, dass die Vergleichsprozesse lediglich als eine mögliche Erklärung für die beobachteten Effekte herangezogen werden können, wird damit überwunden.

Die Befunde beider Studien zeigen, dass dimensionale Vergleichsinformation zu Veränderungen im Fähigkeitsselbstkonzept führt. In Übereinstimmung mit den Annahmen des I/E-Modells (Marsh, 1986) führen – wie beide Studien zeigen – dimensionale Abwärtsvergleiche (Vergleiche der Leistung in einer Aufgabe mit schlechteren Leistungen in einer anderen Aufgabe) zur Erhöhung von Fähigkeitsselbstkonzepten. Dieses Ergebnis repliziert die Befunde von Möller und Köller, 2001a. Darüber hinaus konnte in Studie 2 – unseres Wissens erstmalig – experimentell gezeigt werden, dass dimensionale Aufwärtsvergleiche zu einem Absinken des Fähigkeitsselbstkonzepts führen. Die rückgemeldete Leistung in einer Vergleichsaufgabe wird also als Referenz für die Beurteilung der eigenen Leistungsstärke in der Bezugsaufgabe herangezogen, um die eigenen Fähigkeiten zu erschließen. Bei diesen Effekten handelt es sich um reine Effekte des

Leistungsvergleichs. Da Leistungsunterschiede zwischen den Experimentalgruppen in beiden Studien nicht bestehen, können die Effekte der Rückmeldung nicht auf Leistungsunterschiede zurückgeführt werden. In Studie 2 konnten auch etwaige Unterschiede im allgemeinen Fähigkeitsselbstkonzept zwischen den Experimentalgruppen als Erklärung für die Effekte ausgeschaltet werden. Die zentrale Annahme von Bezugsrahmentheorien (vgl. Skaalvik & Skaalvik, 2002), dass erst der unter Nutzung von Referenzrahmen vorgenommene Leistungsvergleich Effekte auf das Fähigkeitsselbstkonzept hat, wird durch die Ergebnisse gestützt. Es ist erstaunlich, dass wir mit unseren Studien trotz zeitlich nah beieinander liegenden Messungen der Selbstkonzepte differentielle Veränderungen je nach Rückmeldung nachweisen können.

Die Befunde aus Studie 2 zeigen, dass relative Misserfolge in einem Vergleichsbereich zu kompensierenden Aufwertungen im Fähigkeitsselbstkonzept in Bezug auf den Zielbereich führen und analog dazu relative Erfolge zu kompensierenden Abwertungen. Zu beachten ist hierbei die scheinbare Verschiedenheit der Aufgaben aus dem Vergleichsbereich ("Wortanalogien") und dem Zielbereich ("Figurenanalogien"), die von den Versuchspersonen freien Äußerungen zufolge als der verbalen und der mathematischen Domäne zugehörig empfunden wurden. Neuere Arbeiten zu dimensionalen Vergleichen im Schulkontext weisen darauf hin, dass sich Kompensationseffekte nicht zwischen allen schulischen Fächern andeuten, sondern dass es zwischen bestimmten Fächern Assimilationseffekte zu geben scheint (so fanden Schilling, Sparfeldt & Rost, im Druck, sowie Streblov, 2004, positive Effekte der Leistung in Physik auf das mathematische Fähigkeitsselbstkonzept und der Leistung in Mathematik auf das

physikalische Fähigkeitsselbstkonzept). Auch unter praktischen Gesichtspunkten wäre es interessant zu untersuchen, ab welchem Maß an Ähnlichkeit von Aufgaben oder Fächern es im Fähigkeitsselbstkonzept des Zielbereichs nicht zur Kompensation (in Folge von Kontrastierung), sondern zu Generalisierung (in Folge von Assimilation) von Misserfolgen bzw. Erfolgen kommt.

Die Effekte sind in unseren beiden Studien jedoch (ebenso wie in den Studien von Möller, 2000 und Möller & Köller, 2001a) auf das Fähigkeitsselbstkonzept in der Referenzaufgabe A beschränkt. Um zu erreichen, dass die Versuchspersonen die manipulierte Rückmeldung tatsächlich aus der Perspektive von Aufgabe A wahrnehmen, wurde die Aufgabe „Figurenanalogien“ als Bezugsaufgabe in den Fokus der Versuchspersonen gerückt, während die "Wortanalogien" lediglich als zu Vergleichszwecken zu bearbeitende Aufgabe vorgestellt wurden. Weiterhin ist die Richtung der Rückmeldung festgelegt: Die Versuchspersonen erhalten Information darüber, wie sie in der Aufgabe A (Figurenanalogien) im Vergleich zu Aufgabe B (Wortanalogien) abgeschnitten haben. Dies hat möglicherweise zur Folge, dass die Rückmeldungen aus der Perspektive der Aufgabe Figurenanalogien interpretiert wird und aus ihr deshalb weniger stark Information über die eigene Leistungsfähigkeit in Bezug auf die verbalen Aufgaben (Wortanalogien) entnommen wird. Eine andere denkbare Erklärung für diesen Befund liefert die Vermutung von Möller und Köller (2001b), der zufolge verbale Fähigkeitsselbsteinschätzungen - wegen der Vielfalt der Möglichkeiten, im Alltag sprachbezogene Leistungserfahrungen zu machen - möglicherweise durch Leistungsrückmeldungen (z.B. in der Schule oder im Experiment) weniger gut beeinflusst werden können. In Folgestudien können diese konkurrierenden Vermutungen getestet

werden, indem nicht die Figurenalogien als Referenzaufgabe verwendet werden, sondern die Wortanalogien.

Einschränkend ist bei der Interpretation unserer Befunde zu beachten, dass die Manipulation der Leistungsrückmeldung in beiden Studien einen direkten dimensional Vergleich beinhaltete, dieser also nicht spontan von den Versuchspersonen vorgenommen wurde. Dass die alleinige Vorgabe des Vergleichs die Effekte produziert, ist jedoch äußerst unwahrscheinlich. So konnten Möller und Köller (2001 a, Studie 1) die Bedeutung dimensionaler Kontrastierungen auch dann nachweisen, wenn nicht ausdrücklich bereits in der Leistungsrückmeldung dimensional verglichen wurde.

Hinsichtlich des Fähigkeitsselbstkonzepts "Wortanalogien" zeigten sich in beiden Studien unerwartete Haupteffekte. Der beobachtete Haupteffekt der Rückmeldung auf das Fähigkeitsselbstkonzept Wortanalogien in Studie 1 verweist darauf, dass hinsichtlich des Selbstkonzepts Wortanalogien die Gleichverteilung auf die Rückmeldebedingungen nicht gelungen ist. Der sich ergebende Effekt des Messwiederholungsfaktors ist schwer zu erklären. Insbesondere weil die Schwierigkeit der Beispielaufgaben parallel zu den zu bearbeitenden Aufgaben gestaltet wurden und der Effekt der Messwiederholung nur beim Fähigkeitsselbstkonzepts "Wortanalogien", nicht aber bei "Figurenalogien" auftrat, ist es nicht stichhaltig, den Abfall im Fähigkeitsselbstkonzepts Wortanalogien dadurch zu erklären, dass die zu bearbeitenden Aufgaben als gegenüber den Beispielaufgaben unterschiedlich schwer empfunden wurden. Da sich in den vorliegenden beiden Studien infolge der Rückmeldung keine differentiellen Veränderungen im Fähigkeitsselbstkonzept Wort-

analogien ergeben, sind die Befunde für die theoretischen Annahmen ohne Belang.

Entgegen den Befunden von Möller und Köller (2001a) konnten wir in Studie 2 sowohl Veränderungen infolge von dimensionalen Aufwärts- wie Abwärtsvergleichen zeigen. Unsere Befunde sprechen damit gegen die Vermutung von Möller und Köller (2001a), dass dimensionale Aufwärtsvergleiche in experimentellen Settings nicht nachweisbar seien (da im Experiment die Rückmeldesituation wenig bedrohlich für das Selbst der Person ist). Vielmehr ist zu vermuten, dass das Ausbleiben des Effekts infolge dimensionaler Aufwärtsvergleiche in Studie 2 von Möller und Köller (2001a) und der vorliegenden Studie 1 darin begründet liegt, dass der ersten Selbsteinschätzung vergleichsweise leichtere Aufgaben zugrunde lagen als der zu t2 und dies die Effekte dimensionaler Vergleiche überlagert.

Kontrastierende dimensionale Vergleiche führen dazu, dass - wie im Theorieteil erwähnt - fachspezifische Fähigkeits-selbstkonzepte deutlicher disparieren als die entsprechenden Leistungen. Die psychische Funktion dieses Prozesses vermuten Möller und Köller (2004) darin, dass sie Information über die vielfältigen Facetten des eigenen Selbst verstärkt wird. Vor diesem Hintergrund wäre zu vermuten, dass Wissen über eigene Stärken und Schwächen das Treffen von Entscheidungen erleichtert, möglicherweise weil es zu einem besseren Abschätzen der relativen Erfolgsaussichten (vgl. Gollwitzer, 1991) führt. Dies ist eine wichtige Funktion angesichts eines Bildungssystems, welches Individuen an zahlreichen Stellen (etwa bei der Wahl von Grund- und Leistungskursen, bei der Wahl von Vertiefungsfächern oder Ausbildungsberufen) die Möglichkeit eröff-

net, je nach eigenem Profil und eigenen Interessenschwerpunkten den weiteren Weg einzuschlagen.

## Literaturverzeichnis

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

Dickhäuser, O. & Galfe, E. (2004). Besser als..., schlechter als... - Leistungsbezogene Vergleichsprozesse in der Grundschule. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36, 1-9.

Dickhäuser, O., Schöne, C., Spinath, B. & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept: Konstruktion und Überprüfung eines neuen Instrumentes. Zeitschrift für Differentielle und Diagnostische Psychologie, 23, 393-405.

Gollwitzer, P. M. (1991). Abwägen und Planen. Göttingen: Hogrefe.

Heller, K., Gädicke, A.K. & Weinläder, H. (1985). Kognitiver Fähigkeitstest für 4. bis 13. Klassen. Weinheim: Beltz.

Lüdtke, O., Köller, O., Artelt, C., Stanat, P. & Baumert, J. (2002). Eine Überprüfung von Modellen zur Genese akademischer Selbstkonzepte: Ergebnisse aus der PISA-Studie. Zeitschrift für Pädagogische Psychologie, 16, 151-164.

Marsh, H.W. (1986). Verbal and math self-concepts: an internal/external frame of reference model. American Educational Research Journal, 23, 129-149.

Marsh, H.W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The self description questionnaire. Journal of Educational Psychology, 86, 439-456.

Meyer, W.-U. (1984). Das Konzept von der eigenen Begabung. Bern: Hans Huber.

Möller, J. (2000). Effekte dimensionaler und sozialer Vergleiche auf Fähigkeitseinschätzungen und die Zufriedenheit mit der Leistung. Zeitschrift für Experimentelle Psychologie, 47, 67-71.

Möller, J. & Köller, O. (1998). Dimensionale und soziale Vergleiche nach schulischen Leistungen. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 30, 118-127.

Möller, J. & Köller, O. (2001a). Dimensional comparison: An experimental approach to the internal/external frame of reference model. Journal of Educational Psychology, 93, 826-835.

Möller, J. & Köller, O. (2001b). Frame of reference effects following the announcement of exam results. Contemporary Educational Psychology, 26, 277-287.

Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. Psychologische Rundschau, 55, 19-27.

Möller, J., Pohlmann, B., Streblov, L. & Kauffmann, J. (2002). Die Spezifität von Begabungsüberzeugungen als Determinante des verbalen und mathematischen Begabungsselbstkonzepts. Zeitschrift für Pädagogische Psychologie, 16, 87-97.

Rost, D.H., Dickhäuser, O., Sparfeldt, J.R. & Schilling, S.R. (2004). Fachspezifische Selbstkonzepte und Schulleistungen im dimensional Vergleich - Eine versuchsplanerische Überprüfung des I/E-Modells. Zeitschrift für Pädagogische Psychologie, 18, 43-52.

Schilling, S. R., Sparfeldt, J. R. & Rost, D.H. (im Druck). Wie generell ist das Modell? Analysen zum Geltungsbereich des Internal/External Frame of Reference-Modells. Zeitschrift für Pädagogische Psychologie.

Streblow, L. (2004). Bezugsrahmen und Selbstkonzeptgenese (Reihe Pädagogische Psychologie und Entwicklungspsychologie, Band 42). Münster: Waxmann.

Skaalvik, E. M. & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. Educational Psychologist, 37, 233-244.

Wigfield, A. & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement situations. Educational Psychologist, 26, 233-261.

## Fußnoten

<sup>1</sup>Als eine weitere unabhängige Variable wurde Darbietungs- bzw. Bearbeitungsreihenfolge variiert. Während die Hälfte der Versuchspersonen zunächst die Aufgaben des Typs A bearbeitete, beschäftigte sich die andere Hälfte der Versuchspersonen zunächst mit den Aufgaben des Typs B. Dieser Faktor diente zu Kontrollzwecken. Es zeigten sich keine Einflüsse der Reihenfolgevariationen auf die Fähigkeitsselbstkonzepte A bzw. B sowie keine Interaktionen (alle  $F$ -Werte  $< 3.6$ ). Weiterhin wurde der Einfluss eines weiteren experimentellen Faktors exploriert. Um zu erkunden, ob die Effekte dimensionaler Vergleiche im Angesicht einer zu treffenden Wahlentscheidung verstärkt sind (weil die durch dimensionale Vergleiche gelieferte Distinktheitsinformation die Entscheidung erleichtert), wurde beim zweiten Testtermin der einen Hälfte der Vpn angekündigt, sie hätten sich für einen nachfolgenden Wettbewerb gleich für eine der beiden Aufgabenarten A und B zu entscheiden. Der anderen Gruppe der Vpn wurde ebenfalls ein Wettbewerb, jedoch ohne Wahlmöglichkeit zwischen den Aufgaben, angekündigt. Die Effekte dimensionaler Vergleiche fielen im Angesicht der zu treffenden Wahlentscheidung nicht stärker aus als in dem Fall, in dem keine Wahlmöglichkeit bestand,  $F < 0.5$ .

<sup>2</sup>In einem Vorversuch wurden die Vpn im Anschluss an ihre Teilnahme zu dem Versuch befragt. Keine der Versuchspersonen durchschaute das Design oder bezweifelte die Glaubhaftigkeit der Rückmeldung.

<sup>3</sup>Als Beispielaufgaben wurden die Beispielaufgaben für die entsprechenden Tests aus dem KFT (Heller et al., 1985) verwendet. Als Aufgaben für den Test Figurenalogien wurden die Items 41-65 des Subtests N2 des KFT 4-13, Form A verwendet.

Der Test Wortanalogien setzt sich zusammen aus Aufgabe 28-52 des Subtests V4, Form A (Möller, persönl. Mitteilung, 24.2.2004). Die Beispielaufgaben im KFT dienen der Illustration des Prinzips der zu lösenden Aufgaben. Sie sind weniger anspruchsvoll als die Aufgaben selbst. Dies erkennt man unter anderem daran, dass die Beispielaufgaben auch bei Schülern der vierten Klasse verwendet werden, während die verwendeten Testaufgaben für höhere Klassenstufen vorgesehen sind.

<sup>4</sup>Wie bereits in Studie 1 wurde auch in Studie 2 der Einfluss der Ankündigung von Wahlmöglichkeiten überprüft. Die Effekte dimensionaler Vergleiche fielen jedoch auch hier im Angesicht der zu treffenden Wahlentscheidung nicht stärker aus als in dem Fall, in dem keine Wahlmöglichkeit bestand,  $\underline{F} < 0.71$ .

Tabelle 1

Mittleres Fähigkeitsselbstkonzept (FSK) für die Aufgaben „Figurenanalogien“ und "Wortanalogien" in Abhängigkeit von Messzeitpunkt und dimensional vergleichender Rückmeldung (in Klammern Standardabweichung)

	Dimensional aufwärts		Dimensional abwärts	
	t1	t2	t1	t2
	FSK	2.76 <sup>a</sup>	2.44 <sup>b</sup>	2.79
Figurenanalogien	(0.88)	(0.87)	(1.04)	(0.96)
FSK	3.64	3.69	3.23	3.14
Wortanalogien	(0.98)	(0.98)	(0.91)	(0.89)

Anmerkung. Der theoretische Range der Skala liegt zwischen 1 und 5.

Für direkt nebeneinander liegende Zellen zeigen unterschiedliche hochgestellte Indizes, dass die zugehörigen Werte sich statistisch signifikant voneinander unterscheiden ( $p < .05$ ).

Tabelle 2

Mittleres Fähigkeitsselbstkonzept (FSK) für die Aufgaben „Figurenanalogien“ und "Wortanalogien" in Abhängigkeit von Messzeitpunkt und dimensional vergleichender Rückmeldung (in Klammern Standardabweichung)

	Dimensional aufwärts		Dimensional abwärts	
	t1	t2	t1	t2
	FSK	3.91 <sup>a</sup>	3.57 <sup>b</sup>	3.58 <sup>a</sup>
Figurenanalogien	(1.42)	(1.31)	(1.30)	(1.13)
FSK	5.13 <sup>a</sup>	4.69 <sup>b</sup>	4.91 <sup>a</sup>	4.15 <sup>b</sup>
Wortanalogien	(0.90)	(1.08)	(1.07)	(0.86)

Anmerkung. Der theoretische Range der Skala liegt zwischen 1 und 7.

Für direkt nebeneinander liegende Zellen zeigen unterschiedliche hochgestellte Indizes, dass die zugehörigen Werte sich statistisch signifikant voneinander unterscheiden ( $p < .05$ ).