

Soziale Vergleiche interessieren Forschende aus verschiedenen Fachdisziplinen der Sozialwissenschaften seit etwa Mitte des vergangenen Jahrhunderts. Allgemein anerkannt ist, dass seit der Publikation von Leon Festinger im Jahre 1954 soziale Vergleichstheorien als offizielles Fachgebiet gelten (z. B. Blanton & Stapel, 2007; Wheeler & Suls, 2007). Daher gibt es kaum eine einschlägige Veröffentlichung, welche nicht Bezug auf diese wegweisende Publikation von Festinger nimmt.

Einige der zentralen Hypothesen aus Festingers Werk (1954) haben sich bis heute in Theorien bezüglich sozialer Vergleiche durchgesetzt, obschon sie als solche häufig nicht explizit gekennzeichnet sind. Die grundlegendste ist jene, wonach laut Festinger jedem Individuum ein Antrieb inne ist, der zur Evaluation seiner Meinungen und Fähigkeiten anregt, wobei aufgrund vielfach fehlender objektiver Basis Selbstevaluation nur mittels Vergleich mit anderen Personen vorgenommen werden kann. Den Hauptunterschied zwischen Vergleichsprozessen in Bezug auf Meinungen oder Fähigkeiten sah Festinger in der Präferenz von Aufwärtsvergleichen im Falle von Fähigkeiten. Demnach sind soziale Vergleiche tendenziell auf Personen mit subjektiv eingeschätzt höheren Fähigkeiten gerichtet. Festinger unterließ dabei wohlgerne, dass mit dieser Einschätzung bereits eine entsprechende vergleichende Gegenüberstellung vorangegangen sein muss.

Obwohl seine Theorie sozialer Vergleichsprozesse inzwischen als überholt eingestuft und mehrfach kritisiert, weiterentwickelt oder verändert wurde, so dienten doch mehrere ihrer Grundgedanken als Inspiration für etliche bezugsgruppentheoretische Erklärungsansätze und Untersuchungen. In diesem Sinne stellt sie ein „Meisterstück der Mehrdeutigkeit“ dar (Arrowood, 1986, S. 279).

Herbert W. Marsh war einer jener, welcher sich von Festingers Hypothesensammlung inspirieren ließ. Sein sich unter anderem darauf berufender bezugsgruppentheoretischer Erklärungsansatz erhielt bereits in den 1990er-Jahren im pädagogisch-psychologischen Bereich große Beachtung (Abschnitt 2.1). Den prominenten Stellenwert verdankt der von Marsh postulierte *big-fish-little-pond*

effect – oder zu Deutsch Fischteich-Effekt – seinem theoretischen Modell, welches die Wirkungsweise der Referenz- oder Bezugsgruppe auf das Fähigkeitsselbstkonzept zu beschreiben, zu erklären und vorherzusagen vermag. Das Interesse am Konstrukt Selbstkonzept wiederum kann in seiner Beziehung zu Leistung verortet werden. Insofern ist wenig verwunderlich, dass das Selbstkonzept im schulischen Kontext zu einem vieldiskutierten und eingehend erforschten Themengebiet avancierte (Abschnitt 2.2).

Besonderes eindrücklich zeigt sich die Bedeutsamkeit des Selbstkonzepts in sonderpädagogisch ausgerichteten Untersuchungen – bei denen im Zusammenhang mit der Frage nach der angemessenen Bezugsgruppe für Schülerinnen und Schüler mit (sehr) schwachen schulischen Leistungen ebenfalls sein Konfliktpotenzial offenbar wird. Entsprechende Studien und die damit einhergegangenen Debatten werden daher eigens in Abschnitt 2.3 behandelt. Um die entsprechenden Befunde in die aktuelle Diskussion und die Situation des Schweizer Bildungssystems einordnen zu können, wird in Abschnitt 2.4 die Bildung von schulischen Bezugsgruppen und ihre Folgen für die weitere Bildungs- und Berufslaufbahn skizziert, bevor in Abschnitt 2.5 die neuesten Erkenntnisse zum referenzgruppentheoretischen Modell von Marsh sowie die wichtigsten Kritikpunkte aufgezeigt werden, welche bisher an dieses herangetragen wurden. Zuerst folgt nun ein Einblick in die Entstehungsgeschichte des Modells zum *big-fish-little-pond effect*, welche aufschlussreiche Anhaltspunkte für die weitere Auseinandersetzung bietet.

2.1 Referenzgruppentheoretisches Modell nach Marsh

Bezugsgruppentheoretische Überlegungen zur Erklärung von empirisch-pädagogischen Sachverhalten wurden erstmals von James A. Davis im Jahre 1966 dargelegt (J. W. L. Wagner, 1999). Beweggrund war das von ihm als rätselhaft bezeichnete Ergebnis, dass bei amerikanischen College-Studenten die Wahl eines anspruchsvolleren Berufes in einem engeren korrelativen Zusammenhang zum individuellen Notendurchschnitt stand als zur Qualität des Colleges, gemessen an der proportionalen Vertretung fähiger Studierenden im Campus. Die Interpretation seiner Daten resultiert in einem – von Davis selbst als zwar nicht perfekten, aber auch nicht trivial bezeichneten – Rat an Eltern, bei der Auswahl eines „feinen“, hochselektiven Colleges neben den Vorteilen auch die Nachteile zu beachten, da dort die größere Wahrscheinlichkeit bestehe, in den unteren Rängen einer Klasse zu landen – was mit der Verwendung des englischsprachigen Aphorismus „It is better to be a big frog in a small pond than a small frog in a big pond“ (Davis, 1966,

S. 31) so pointiert umschrieben wurde, dass in späteren Arbeiten wiederholt darauf zurückgegriffen (u. a. J. W. Meyer, 1970) oder daran angeknüpft wurde.

Wohl unschwer lässt sich im obigen Gedankensplitter die Fischteich-Metapher von Herbert W. Marsh erkennen. Während in Festingers Hypothesensammlung mögliche Ergebnisse der Vergleichsprozesse weitgehend unberücksichtigt oder zumindest nur sehr vage blieben (vgl. dazu die Kritik von Stapel & Blanton, 2007), fokussierte Marsh explizit auf diese Effekte sowie den bestimmten Referenzrahmen respektive Vergleichsstandard, den eine Gruppe bietet (Marsh, 1987). Das theoretische Modell wurde wie folgt begründet und beschrieben: Basierend auf der Annahme, dass sich Schülerinnen und Schüler aufgrund von Selektionsprozessen im Allgemeinen in einer sozialen Bezugsgruppe befinden, in der die durchschnittliche Schulleistungsfähigkeit der Mitschülerinnen und Mitschüler systematisch von derjenigen einer größeren, repräsentativen Normgruppe abweicht, entwickelt beispielsweise eine Schülerin durch den kontinuierlichen Vergleich mit ihren Klassenkameradinnen und Klassenkameraden innerhalb einer leistungsschwachen Schulklasse eine höhere Einschätzung der eigenen schulischen Leistungsfähigkeit – also ein höheres akademisches Selbstkonzept – als in einer Schulklasse mit einem hohen Leistungsniveau. Nach dem theoretischen Modell von Marsh hat somit die mittlere Schulleistung der Schulklasse oder Schule als soziale Referenzgruppe einen negativen Effekt auf das akademische Selbstkonzept eines einzelnen Schülers oder einer einzelnen Schülerin (siehe Abbildung 1).

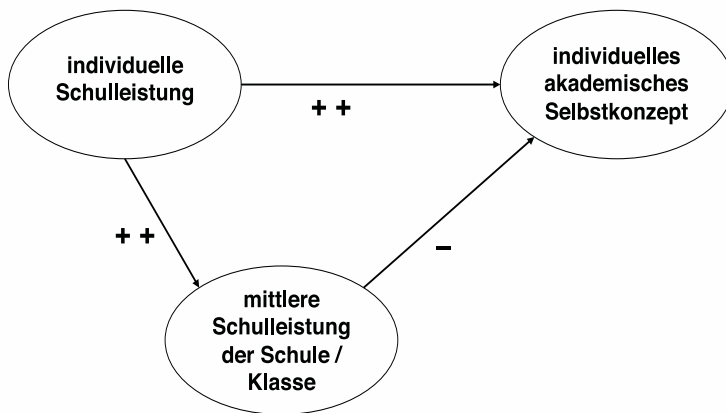


Abb. 1 Klassisches Modell zum *big-fish-little-pond effect* (Marsh, 2005)

Der *big-fish-little-pond effect* (BFLPE) wurde erstmals im Jahr 1984 von Marsh und Parker als solcher benannt und beschrieben. Wie rund zwei Jahrzehnte zuvor bereits Davis, warnten auch Marsh und Parker aufgrund der Ergebnisse aus ihrer Untersuchung vor einer Platzierung in einer Schule mit erhöhten Leistungsansprüchen.

Bemerkenswert ist, dass die ursprüngliche Absicht der Untersuchung von Marsh und Parker (1984) darin bestand, amerikanische Studien (jene von Soares & Soares, 1969; Trowbridge, 1970, 1972) zu replizieren, welche „paradoxerweise“ eine negative Korrelation zwischen dem durchschnittlichen sozioökonomischen Status der Schülerschaft einer Schule und der Selbsteinschätzung der individuellen Fähigkeiten festgestellt hatten. Aufgrund des in ihrer Untersuchung ermittelten sehr starken Zusammenhanges zwischen der für eine Schule durchschnittlichen Leistung und dem sozioökonomischen Status im Mittel kamen die beiden zu der Schlussfolgerung, dass die früheren Studien den negativen Effekt des Besuches einer hoch-anspruchsvollen Schule (dem großen Teich) auf die Wahrnehmung der akademischen Fähigkeiten unterschätzt hatten, und verfassten daraufhin das theoretische Modell zur Beschreibung des *big-fish-little-pond*-Effekts (Marsh, 1987). Die Idee, dass eine angesehene Schule sich positiv auf die Wahrnehmung des einzelnen Schülers auswirken könnte – im Sinne eines *reflected-glory*-Effekts, wurde damals weitgehend verworfen.

In Vergessenheit geriet der reflektierende Ruhmeseffekt – oder auch Assimilationseffekt – dann doch nicht: Rund fünfzehn Jahre später nahmen sich Marsh, Kong und Hau (2000) der Thematik erneut an und untersuchten in Hongkong das dortige, rigide selektionierende Schulsystem. Eingebettet in eine kollektivistische und gleichzeitig enorm leistungsorientierte Gesellschaft, in welcher der Besuch einer renommierten Schule mit einer sozialen Statuserhöhung verbunden ist, sollte mittels eines natürlichen Experiments untersucht werden, ob das hoch-selektive Schulsystem die negative Vergleichswirkung erhöhen oder aber die gesellschaftlichen Bedingungen diesen Effekt vermindern respektive den Ruhmeseffekt verstärken würden. Schülerinnen und Schüler, welche ihrer Schule einen hohen Status zuschrieben, tendierten zu höheren Einschätzungen ihrer akademischen Fähigkeiten, was Marsh und seine Mitarbeiter zum Fazit veranlasste, dass der Besuch einer angesehenen Schule gleichzeitig mit einem anspruchsvolleren Vergleichsstandard und einer Quelle des Ruhmes verbunden sei. Den BFLPE beschrieben sie demgemäß als Nettoergebnis von zwei gegenläufigen Effekten: dem stärkeren negativen Kontrasteffekt des Leistungsdurchschnitts der Schule und dem positiven Assimilationseffekt des wahrgenommenen Status der Schule (ebd.). Der negativ wirkende Vergleich einer Schülerin oder eines Schülers mit anderen leistungstärkeren Mitschülerinnen und Mitschülern schien abgedämpft zu werden von Gedanken wie in etwa: Wenn ich in dieser Schule bin, kann ich gar nicht (so) leistungsschwach sein.

Damit war und ist die Diskussion darüber, ob – und falls ja wie – der sozioökonomische Status einer Schülerin oder eines Schülers sowie der durchschnittliche sozioökonomische Status einer Schule das akademische Selbstkonzept beeinflussen, noch nicht abschließend beantwortet (siehe Abschnitt 2.4.3).

2.2 Selbstkonzept und schulische Leistungsentwicklung

Spätestens jetzt stellt sich die berechtigte Frage, warum eine positive Einschätzung der eigenen schulischen Fähigkeiten wichtig für die schulische Laufbahn sein soll. Ein gutes Selbstwertgefühl könne doch höchstens persönliches Vergnügen bereiten, wenn nicht sogar abträglich für die Leistungsentwicklung sein – meinten etwa Baumeister und Mitarbeitende (2003).

Dem ist grundsätzlich entgegen zu halten, dass klar unterschieden werden muss zwischen dem globalen Selbstwertgefühl (*self-esteem*) und dem Selbstkonzept (*self-concept*), welches sich außerdem in verschiedene spezifische Komponente unterteilen lässt (u. a. Marsh & Craven, 2006; Moschner & Dickhäuser, 2006; Trautwein, Lüdtke, Köller, & Baumert, 2006; Wigfield & Wagner, 2007).

2.2.1 Akademisches Selbstkonzept

Ganz allgemein betrachtet kann das Selbstkonzept als eine Gedächtnisstruktur definiert werden, welche Informationen über sich selbst beinhaltet. Um beispielsweise eine Einschätzung über die eigenen Kompetenzen im Lesen abgeben zu können, wird auf jenes Wissen über sich selbst zurückgegriffen, welches über etliche Situationen hinweg und in Interaktion mit anderen Personen gesammelt wurde (Wild, Hofer, & Pekrun, 2006).

Mit dem Begriff „Gedächtnisstruktur“ wird bei dieser Definition des Selbstkonzepts die *kognitive* Repräsentation der eigenen *spezifischen* Fähigkeiten betont. Im Gegensatz stellt das Konstrukt Selbstwertgefühl eine *generelle* und *affektiv* besetzte Selbstbewertungstendenz dar (u. a. Jerusalem & Schwarzer, 1991; Schöne, Dickhäuser, Spinath, & Stiensmeier-Pelster, 2003; Wigfield & Wagner, 2007). Die beiden Unterscheidungsmerkmale – spezifisch versus generell und kognitiv versus affektiv – mögen trivial erscheinen. Mit deren Ignorierung lassen sich jedoch einige der Diskrepanzen bei Befunden aus den Anfängen der Selbstkonzept-Forschung erklären.

Mit der nahezu inflationären Verwendung des Begriffs Selbstkonzept wurde damals – und wird teilweise auch heute noch – vorausgesetzt, dass mehr oder weniger „jeder weiß, um was es dabei geht“; was häufig dazu führte, dass weder eine fundierte theoretische Definition noch eine systematische Überprüfung der eingesetzten Messinstrumente vorhanden waren (Marsh, 1993). Die Periode zwischen 1960 und den frühen 1980er-Jahren wurde von John Hattie (1992) denn gar als die Blütezeit der *Dust Bowls* in der Selbstkonzept-Forschung betitelt – in Anlehnung an die heftigen Staubstürme in den großen Ebenen von Nordamerika, welche während der Weltwirtschaftskrise in den 1930ern für die Urbarmachung gerodet wurden.

Ein Fehlen entsprechender theoretischer Modelle wurde schon in den 1970er-Jahren bemängelt (u. a. Shavelson, Hubner, & Stanton, 1976; Wells & Marwell, 1976). Wegweisend für die weitere Forschung zum Konstrukt Selbstkonzept war in erster Linie die systematische Überblicksarbeit von Richard J. Shavelson und seinen Mitarbeitenden (siehe dazu die Reviews von Byrne, 1996; Hansford & Hattie, 1982; Marsh & Hattie, 1996). Basierend auf sieben charakteristischen Merkmalen für eine angemessene Definition des Konstrukts Selbstkonzept – organisiert, multidimensional, hierarchisch, stabil versus instabil (bzw. generell versus situativ), entwicklungsorientiert, deskriptiv und evaluativ sowie unterscheidbar von anderen Konstrukten – entwickelten Shavelson und Mitarbeitende zur Veranschaulichung der Konzeptualisierung ein *multidimensionales, hierarchisch-organisiertes Modell* (siehe Abbildung 2). Nach diesem Modell kann das *generelle Selbstkonzept* auf einer ersten Ebene unterteilt werden in eine *akademische* und eine *nicht-akademische Dimension*. Das akademische Selbstkonzept lässt sich wiederum aufgliedern anhand verschiedener *Teilbereiche* wie Mathematik, Geschichte oder Sprache. Auf einer weiteren Ebene befinden sich dann die einzelnen spezifischen *Situationen*, in denen das jeweilige Verhalten zum Ausdruck kommt und von der betreffenden Person zur Selbstevaluation herangezogen wird.

Obwohl ursprünglich nur als eine mögliche Repräsentation der hierarchischen Eigenschaft ihrer Selbstkonzept-Definition gedacht, wird diese seither als *das Shavelson-Modell* bezeichnet (u. a. Marsh & Hattie, 1996), welches mehrfach empirisch überprüft und weiterentwickelt wurde (u. a. Byrne, 1986, 1996; Shavelson & Bolus, 1982).

Einhergehend mit der Entwicklung der Strukturgleichungsmodellierung wurde mittels konfirmatorischer Faktorenanalysen die hierarchische Ordnung der verschiedenen Teilkomponenten untersucht. Die Resultate waren konsistent mit der Annahme, dass das Selbstkonzept hierarchisch geordnet ist. Die Strukturierung der ersten, höheren Ebene stellte sich hingegen als komplexer als ursprünglich angenommen heraus. Anstelle eines einzelnen akademischen Faktors wurden deren zwei festgestellt: das *mathematische akademische* und das *verbale akademische Selbst-*

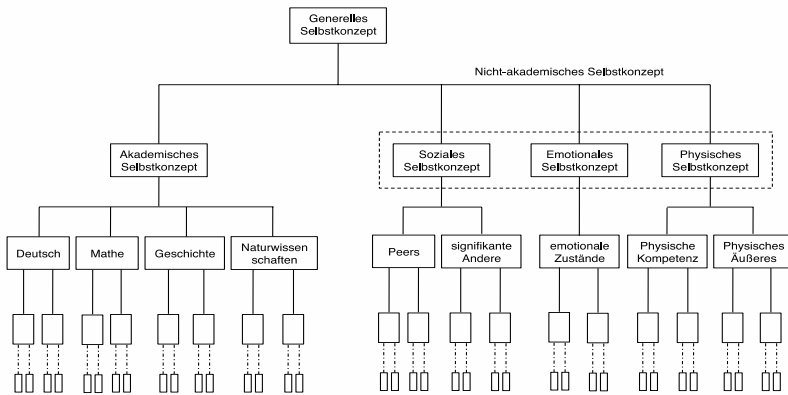


Abb. 2 Repräsentation der hierarchischen Organisation des Selbstkonzepts nach Shavelson, Hubner und Stanton (1976, S. 413)

konzept (Marsh, 1990b; Marsh & Shavelson, 1985; Marsh, Shavelson, & Byrne, 1988). Dieses so genannte *Marsh/Shavelson-Modell* mit seinen zwei akademischen Domänen dominierte in den letzten rund 25 Jahren die Forschungslandschaft rund um das akademische Selbstkonzept. Einzelne Ungereimtheiten forderten einige der zentralen Hypothesen des theoretischen Modells heraus: So entdeckten beispielsweise Marsh und Yeung (2001), dass das Englische und Spanische Selbstkonzept von Schülerinnen und Schülern in den USA nahezu unkorreliert waren (ähnliche Resultate auch bei Marsh, Kong, & Hau, 2001). Bei der Berücksichtigung eines breiteren Spektrums an Schulfächern zeigte sich zudem zwischen dem mathematischen und dem verbalen akademischen Selbstkonzept einen substantiellen Zusammenhang; die beiden Konstrukte teilten außerdem einen großen Anteil der Varianz des generellen akademischen Selbstkonzepts (Marsh, 1990b).

Diese und weitere widersprüchliche Befunde führten Brunner und Mitarbeitende zu einer Überprüfung verschiedener alternativer Modelle zum akademischen Selbstkonzept (Brunner et al., 2010; Brunner, Keller, Hornung, Reichert, & Martin, 2009). Neben dem Shavelson-Modell (1) und dem Marsh/Shavelson-Modell (2) wurden weiter ein strukturelles Modell erster Ordnung mit verschiedenen fachspezifischen Selbstkonzepten (3) sowie ein Modell mit einem generellen akademischen Selbstkonzept als übergeordneten Faktor (4) mittels konfirmatorischer Faktorenanalysen analysiert und einem Modellvergleich unterzogen. Die Ergebnisse unterstützen das vierte Modell, bei dem ein *generelles* akademisches Selbstkonzept den Apex der

Hierarchie über die verschiedenen domänenspezifischen Selbstkonzepte einnimmt (siehe Abbildung 3).

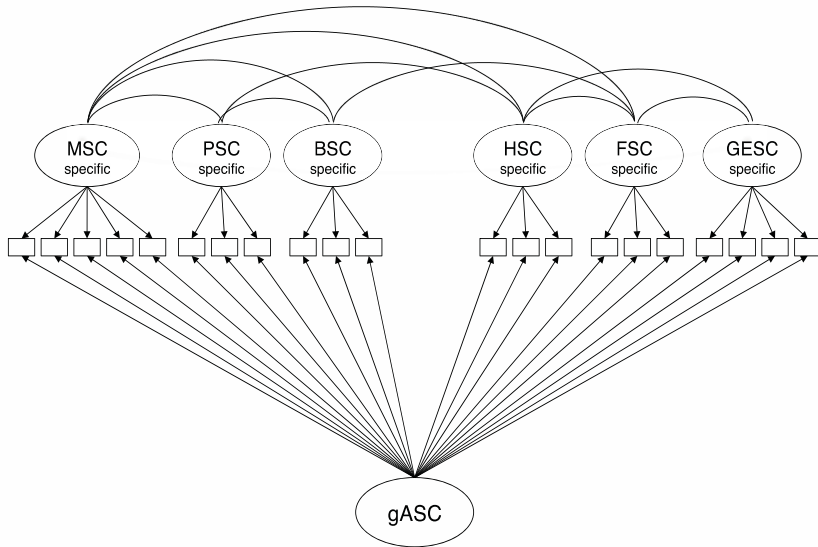


Abb. 3 Strukturelle Konzeption des akademischen Selbstkonzepts gemäß dem *nested Marsh/Shavelson model* (Brunner et al., 2010, S. 966)

Anmerkung. gASC: generelles akademisches Selbstkonzept, specific: spezifisches Selbstkonzept in einem bestimmten Fachbereich (z. B. MSC specific: spezifisches mathematisches Selbstkonzept)

Nach diesen neueren empirischen Erkenntnissen wird das akademische Selbstkonzept als fachspezifisch und hierarchisch organisiert verstanden. Gegenüber dem Marsh/Shavelson-Modell wird allerdings auf der zweiten, übergeordneten Ebene nicht zwischen einem generellen mathematischen und einem generellen verbalen akademischem Selbstkonzept unterschieden – was einer Revision des Marsh/Shavelson-Modells entspricht. Das hierarchische Modell, mit einem generellen Faktor und mit je nach Domäne distinktiven Selbstkonzepten, wird als *nested Marsh/Shavelson model* respektive ineinander verschachteltes Marsh/Shavelson-Modell bezeichnet (Brunner et al., 2010).

2.2.2 Genese und Entwicklung des akademischen Selbstkonzepts

In very broad terms, self-concept is a person's perception of himself. These perceptions are formed through his experience with his environment ... and are influenced especially by environmental reinforcement and significant others. We do not claim an entity within a person called "self-concept". Rather, we claim that the construct is potentially important and useful in explaining and predicting how one act. (Shavelson et al., 1976, p. 411)

Beim Selbstkonzept einer Person handelt es sich gemäß Definition von Shavelson und Mitarbeitenden keineswegs um eine rein individuelle Angelegenheit. Es wird geformt durch ihre Erfahrungen mit dem Umfeld und beeinflusst durch die jeweiligen Bezugspersonen. Die Entwicklung des akademischen Selbstkonzepts wird somit in erheblichem Maße durch die spezifischen Bedingungen des schulischen Umfeldes, den Mitschülerinnen und Mitschülern sowie den Lehrpersonen als relevante Bezugspersonen beeinflusst (u. a. Harter, 1998; Jerusalem & Schwarzer, 1991).

Wie entsteht und entwickelt sich nun bei einem Kind oder Jugendlichen ein solches Konzept über die eigenen schulischen Fähigkeiten? Oder konkreter: Ab wann beginnt sich das akademische Selbstkonzept eines Kindes auszubilden: unmittelbar bei Eintritt in die Schullaufbahn, beim Einsetzen von Leistungsbeurteilungen oder erst später z. B. in der frühen Adoleszenz? Und vor allem: Ab wann ziehen Kinder oder Jugendliche Leistungen anderer aus der schulischen Bezugsgruppe zum Vergleich heran?

Selbstkonzeptentwicklung während der Kindheit

Laut einem Forschungsüberblick von Ruble und Frey (1991) werden bis zum Alter von sieben bis acht Jahren nur selten soziale Vergleiche zur Selbsteinschätzung von Kompetenzen herangezogen. Dies bedeutet allerdings nicht, dass jüngere Kinder sich noch nicht mit Gleichaltrigen vergleichen. Vorschulkinder zeigen etwa deutliches Interesse an spielerischen Wettkämpfen mit den anderen Kindern aus der vorschulischen Gruppe (Butler, 1989a, 1989b), scheinen bereits eine einigermaßen angemessene Vorstellung der Rangordnung von gewissen (z. B. sportlichen) Leistungen zu haben (Morris & Nemcek, 1982) und äußern spontan soziale Vergleiche insbesondere im Sinne von Ähnlichkeiten und Unterschieden (z. B. „Mein Schiff ist gleich groß wie deins“) oder im kompetitiven Sinne (z. B. „Ich kann schneller rennen als du“) (Mosatche & Bragonier, 1981). Kinder im Vorschulalter benutzen den Vergleich mit anderen Kindern also nicht zwecks Selbstevaluation sondern noch vornehmlich mit dem Bestreben *konkrete* Informationen über Ähnlichkeiten und Unterschiede herauszufinden. Erst in den darauffolgenden Schuljahren beschäftigen

sich Kinder zunehmend damit, *wie* sie beispielsweise in Schularbeiten im Vergleich zu den Mitschülerinnen und Mitschülern abschneiden, oder wie sie Aufgaben im Vergleich zu den anderen erledigen (Schunk, Pintrich, & Meece, 2010).

Laut Altermatt und Kolleginnen (2002) lässt sich bereits im Kindergarten und während den ersten beiden Schuljahren eine Beeinflussung der Sichtweise der eigenen Kompetenzen durch die Art und Weise der Äußerungen von Klassenkolleginnen und -kollegen beobachten. Bedeutsam für die Herausbildung des Selbstkonzepts scheinen auch die jeweiligen pädagogischen Rückmeldepraktiken zu sein: In einer quasi-experimentellen Studie von Stipek und Daniels (1988) zeigten Kinder aus Kindergartenklassen, in denen normative Beurteilungen explizit hervorgehoben wurden, ein deutlich tieferes Selbstkonzept als jene aus Kindergartenklassen, in denen kaum Leistungsrückmeldungen abgegeben wurden. Die Ausprägung des Selbstkonzepts von Kindern aus den Kindergartenklassen mit Beurteilung war zudem vergleichbar mit dem Selbstkonzept von Schülerinnen und Schüler der vierten Grundschulklasse. Die Autorinnen schlossen daraus, „that typical changes in classroom environments and the nature of instruction over the elementary school years explain, in part, the decline found in most developmental studies of children’s perceptions of academic competence“ (ebd., p. 355).

Eine der typischen Veränderungen von der Vorschule zur Primarstufe ist die Einführung von Schulnoten. Im Rahmen eines Schulversuchs des Deutschen Bundeslandes Bayern wurden Kinder aus Schulklassen der zweiten Jahrgangsstufe mit Notengebung und solche aus Schulklassen ohne Notengebung verglichen (Zeinz, 2006). Die Ergebnisse verwiesen auf einen verstärkten negativen Referenzgruppeneffekt in den Klassen mit Notengebung, welcher dadurch erklärt wurde, dass „tägliche Vergleiche der Kinder mit ihren Mitschülerinnen und Mitschülern im Unterrichtsgeschehen einen Einfluss auf das mathematische Selbstkonzept nehmen und die Notengebung diesen Einfluss verstärkt“ (Zeinz & Köller, 2006, S. 186).

In einer Stichprobe der Deutschschweizer Schuleingangsstufe, welche Schulklassen der sogenannten Grundstufe (Kindergarten plus 1. Primarklasse), der Basisstufe (Kindergarten plus 1./2. Primarklasse) und der ersten beiden Primarschuljahre umfassten, zeigten sich bereits bei Kindern im Alter zwischen rund sieben und neun Jahren komparative Referenzgruppeneffekte. In Bezug auf die Schulform der Eingangsstufe zeigten sich keine signifikanten Unterschiede (Abt Gürber, 2011).

Weitere Unterstützung, dass soziale Vergleichsprozesse gleich ab Schulbeginn auf die Selbsteinschätzung der schulischen Leistungen wirken, bietet eine Studie von Chapman, Tunmer und Prochnow (2000), welche die Entwicklung des Selbstkonzepts der Lesefähigkeiten bei Schülerinnen und Schülern während den ersten zweieinhalb Schuljahren verfolgt hatten. Bereits nach den ersten sechs bis acht Schulwochen zeigten sich Unterschiede in den Selbsteinschätzung der Lesefähigkeiten sowie in

den Einstellungen gegenüber dem Lesen. Ein tiefes Selbstkonzept bezüglich Lesen hatten generell jene Kinder, die mit wenig Vorkenntnissen bezüglich phonologischer Bewusstheit und kaum Buchstabenkenntnissen in die Schulzeit eingestiegen waren. Mitte des dritten Schuljahres schätzten sich diese Kinder dann als weniger kompetent im Lesen ein als die anderen Kinder und äußerten sich dahingehend, mehr Schwierigkeiten im Lesen und weniger Freude daran zu haben.

Die Ergebnisse von Chapman und seinen Mitarbeitern stehen damit in Widerspruch zu den im deutschsprachigen Raum vielzitierten Studien von Helmke, wonach Schulkinder zu Beginn des dritten Schuljahres im Mittel ein ebenso erhöhtes allgemeines akademisches Selbstkonzept wie im Kindergarten aufwiesen (Helmke, 1991). Während zu Beginn der Grundstufe (entspricht der Primarstufe) die Selbstbeurteilungen im Allgemeinen eher „optimistisch“ und wenig differenziert ausfielen, wurde gegen Ende der Grundstufe eine Angleichung an die „objektiven“ Einschätzungen sowie eine Differenzierung des Selbstkonzepts beobachtet (Helmke, 1998).

Die bereits 1976 von Shavelson, Hubner und Stanton formulierte Differenzierungshypothese, wonach Kinder mit zunehmendem Alter über ein differenzierteres Selbstkonzept im Sinne eines multidimensionalen Modells verfügen, findet zwar auch andernorts Unterstützung (u. a. Marsh & Ayotte, 2003). Mehrheitlich wird allerdings schon fünf- bis sechsjährigen Kindern die Fähigkeit zur Unterscheidung zwischen verschiedenen Bereichen von Kompetenzen zugesprochen (Marsh, Craven, & Debus, 1991; Marsh, Ellis, & Craven, 2002).

Selbstkonzeptentwicklung während der Adoleszenz

Die Adoleszenz wurde lange Zeit als die Phase der größten Veränderungen des Selbstkonzepts betrachtet (Wigfield, Eccles, & Pintrich, 1996). Diese Annahme basiert unter anderem auf dem wegweisenden Werk von Erikson (1968), der die Adoleszenz als Phase der Identitätsbildung charakterisierte: Jugendliche sind demzufolge zwischen ihrem dreizehnten und zwanzigsten Lebensjahr auf der Suche nach ihrer Identität, welche schlussendlich gefunden wird, oder falls nicht, in einer Identitätsdiffusion oder Identitätskrise endet. Konsistent mit Eriksons Stufenmodell der psychosozialen Entwicklung wurde das Selbstkonzept von Jugendlichen in der Folge häufig als weniger stabil und weniger ganzheitlich beschrieben (Harter, 1990, 1999). Daher galt die Integration der verschiedenen Aspekte des Selbst lange Zeit als wichtige Entwicklungsaufgabe der Adoleszenz.

Im Zuge der bereits erwähnten Kritik an zu wenig theoriebasierter Selbstkonzeptforschung und mangelhafter Überprüfung der verwendeten Instrumente (vgl. Abschnitt 2.2.1) fokussieren sich Forschende seither mehrheitlich auf spezifischere Aspekte anstelle der Erfassung eines allgemeinen Verständnisses von sich selbst. Bei der Überprüfung des eigens für die Erfassung des Selbstkonzepts von

Jugendlichen entwickelten Instruments *Self Description Questionnaire III* (SDQ III) von Marsh und O'Neill (1984), zeigte sich, dass das generelle Selbstkonzept von Jugendlichen weniger stark als üblicherweise rapportiert mit den spezifischen Selbstkonzeptbereichen zusammenhing. Als Begründung für diesen Befund geben die Autoren zum einen an, dass das Instrument aufgrund seiner psychometrischen Eigenschaften eine bessere Differenzierung ermöglichen würde als vorangehende Instrumente, und zum anderen, dass die Jugendlichen – im Gegensatz zu jüngeren Kindern – besser zwischen den verschiedenen Facetten des Selbstkonzept unterscheiden könnten (ebd.).

Während in der Kindheit die Entwicklung des Selbstkonzepts sowohl in Richtung zunehmender Differenzierung als auch Hierarchisierung verläuft, prägen sich in der Adoleszenz in erster Linie die einzelnen Bereiche des Selbstkonzepts stärker aus. Das Selbstkonzept scheint in der Jugendzeit demgemäß weniger hierarchisch geordnet, Multidimensionalität bleibt hingegen ein wichtiges Charakteristikum. Gleichzeitig ist während der Vorpubertät und in der frühen Adoleszenz eine Abnahme des Selbstkonzeptes zu beobachten – was vielfach als Herausbildung eines „realistischen“ oder „objektiven“ Selbstkonzepts angesehen wird (u. a. Marsh, 1989).

Damit ist eine zweite Entwicklungstendenz angesprochen: Während bei jüngeren Kindern häufig von einem überhöhten oder idealistischen Selbstkonzept ausgegangen wird, entwickeln Jugendliche im Verlaufe der Adoleszenz ein wirklichkeitsnahes oder tatsächliches Selbstkonzept. Diese Entwicklung wird in Zusammenhang mit der bereits angesprochenen Identitätsbildung betrachtet.

Identitätsbildung meint im schulischen Kontext in erster Linie den Prozess, während dessen ein Jugendlicher eine „angemessenere“ Wahrnehmung der eigenen relativen Kompetenzen aufbaut (Wigfield & Wagner, 2007). Die Herausbildung eines realitätsnahen Selbstkonzepts kollidiert in der Adoleszenz mit einer weiteren wichtigen Entwicklungsaufgabe dieser durch Veränderungen und Umbrüche gekennzeichneten Phase: der Aufbau von selbst-stabilisierenden Prozessen (Skinner & Zimmer-Gembeck, 2007). Gefragt ist also einerseits die fortwährende Anpassung des Selbstkonzepts und andererseits die Stabilisierung beziehungsweise Verteidigung – oder: ein „Kompromiss zwischen Realismus und Stabilisierung“ (Greve et al., 2009, S. 27). Eine Anpassung im Sinne von „realistischen“ Einschätzungen der Möglichkeiten und Grenzen ist für Jugendliche besonders im Hinblick auf die Berufswahl oder weiterführende Ausbildungen von Bedeutung. Je nach Situation werden Stabilisierungsprozesse bis hin zu Verteidigungsreaktionen vorteilhaft oder notwendig.

Bei einem selektiven Bildungssystem wie jenem der Schweiz liegt die Vermutung nahe, dass solche Prozesse bereits beim Übergang von der Primarstufe auf die Sekundarstufe I einsetzen. Gemäß Jerusalem (1993) ist dieser zentrale Angelpunkt

der obligatorischen Schulzeit von einschneidender Bedeutung für die Entwicklung des akademischen Selbstkonzepts. Mit den Auswahlverfahren für die Zuweisung auf die verschiedenen Schultypen der Sekundarstufe I geraten soziale Vergleiche unweigerlich in den Fokus. Des Weiteren werden durch die Leistungsgruppierung auf der Sekundarstufe I, durch die Berufswahl und durch die Auswahlverfahren für weiterführende Bildung soziale Vergleiche betont und Konkurrenzdenken intensiviert (Wigfield et al., 1996). Vor diesem Hintergrund wird verständlich, dass die vorgängig angesprochene Abnahme in der Ausprägung des akademischen Selbstkonzepts während der Adoleszenz im Zusammenhang mit dem schulischen Übergang steht (u. a. Aust, Watermann, & Grube, 2010; Eccles et al., 1993).

„Selbstimmunisierung als realitätsadaptiver Stabilisierungsprozess“ scheint laut Greve und seinen Mitarbeitern (2009, S. 30) daher vor allem zu Beginn der Adoleszenz – oder übertragen auf die Bildungslaufbahn: zu Beginn der Sekundarstufe I – besonders nützlich zu sein. Im Verlaufe der Adoleszenz nimmt diese „Pufferwirkung“ kontinuierlich ab; vermutlich weil mit zunehmendem Alter die soziale Anschlussfähigkeit zusehends wichtiger wird (Arnett, 2000). Die Befunde von Greve und Mitarbeitenden sind jedoch mit Vorsicht zu betrachten, da sie – obwohl mehrfach vom Selbstkonzept die Rede ist – auf der Erfassung des Selbstwerts (nach Rosenberg, 1965) basieren.

Das Thema „realistisches“ Selbstkonzept ist damit noch nicht abschließend behandelt. Gerade in Integrations-Separations-Debatten wird die Argumentation „wirklichkeitsnahes versus wirklichkeitsfremdes“ Selbstkonzept immer wieder eingebracht. Daher wird diese Problematik im Zusammenhang mit Ergebnissen aus der sonderpädagogisch ausgerichteten Forschung in Abschnitt 2.3 erneut aufgegriffen.

2.2.3 Kausalitätsdilemma zu Selbstkonzept und Schulleistung

Kaum ein anderes Konstrukt der Sozialwissenschaften erhält dieselbe universelle und multidisziplinäre Beachtung wie das Selbstkonzept (u. a. Marsh, Craven, & McInerney, 2003). Seine ungebrochene Beliebtheit in den unterschiedlichsten Forschungsfeldern der Sozialwissenschaften gründet in seiner postulierten oder nachgewiesenen Beeinflussung einer Vielzahl wichtiger oder erwünschter Zielvariablen. Im Kontext der Diskussion um Einflussfaktoren für eine positive schulische Leistungsentwicklung nimmt daher das akademische Selbstkonzept unumstritten eine zentrale Stellung ein. Ein positives akademisches Selbstkonzept wird dabei gleichzeitig als Ziel und Hilfsmittel zur Förderung von schulischen Leistungen gehandelt (Marsh & O'Mara, 2008).

Dabei drängt sich die Frage auf, ob ein hohes akademisches Selbstkonzept die schulische Leistungsentwicklung zu beeinflussen vermag. Konkret: Sind gute Schulleistungen der Grund, dass jemand positiv über die eigenen Schulleistungsfähigkeiten denkt, oder ist ein positives Konzept über die eigene Schulleistungsfähigkeit Anstoß für gute Schulleistungen? Diese Frage – in der einschlägigen Literatur als *the-chicken-or-the-egg problem* oder auch Kausalitätsdilemma bekannt – beschäftigt Selbstkonzeptforschende seit über drei Jahrzehnten (Marsh & Köller, 2003).

Grundsätzlich sind vier Antworten auf die Frage vorstellbar: 1. nein, es besteht kein Zusammenhang zwischen dem akademischem Selbstkonzept und der schulischen Leistung; 2. nein, aber die schulische Leistung beeinflusst das spätere akademische Selbstkonzept (und nicht umgekehrt); 3. ja, das akademische Selbstkonzept beeinflusst die nachfolgenden schulischen Leistungen; und 4. ja, aber es besteht eine reziproke Beeinflussung zwischen akademischem Selbstkonzept und schulischer Leistung. Während die erst genannte Variante kaum je vertreten wurde, erhielten die drei anderen Varianten sowohl von empirischer als auch theoretischer Seite her Unterstützung.

Zur Erklärung der kausalen Anordnung wurden über weite Strecken zwei sich rivalisierende Modelle herangezogen: das *self-enhancement model* und das *skill-developmental model* (wie etwa bereits bei Calsyn & Kenny, 1977). Der *self-enhancement* oder Selbstverstärkung-Ansatz betrachtet das akademische Selbstkonzept als erklärende Variable respektive Prädiktor der schulischen Leistung, während umgekehrt die Schulleistung das akademische Selbstkonzept nicht beeinflusst. Demzufolge plädieren Vertreterinnen und Vertreter dieses Modells für eine Steigerung des akademischen Selbstkonzepts, um die schulischen Leistungen zu erhöhen. Empirische Unterstützung für den *Self-Enhancement* oder Ansatz lieferten unter anderem die Längsschnittstudien von Marsh (1990a) sowie Shavelson und Bolus (1982). Im Gegensatz dazu wird beim *skill-developmental* oder Kompetenzentwicklungs-Ansatz das akademische Selbstkonzept als abhängige Variable respektive Kriterium der Schulleistungen aufgefasst. Entsprechende Effekte der vorangehenden Schulleistungen auf das nachfolgende akademische Selbstkonzept konnten Helmke und van Aken (1995) oder Skaalvik und Valas (1999) in ihren Studien mit längsschnittlichem Design aufweisen.

Im Zuge der Integration beider Modelle zum *reciprocal-effects model* ist die „Huhn-oder-Ei“-Debatte mehr oder weniger verebbt. Die Ergebnisse der Untersuchung von Marsh und Yeung (1997a) und nicht zuletzt der umfassende und ausführliche Review von Marsh und Craven (2006) unterstützten recht deutlich die Annahme eines reziproken Zusammenhangs zwischen akademischem Selbstkonzept und Schulleistungen. Nichtsdestotrotz können die bisherigen Forschungsergebnisse kein eindeutiges Entwicklungsmuster der kausalen Anordnung von akademischem

Selbstkonzept und Schulleistungen aufzeigen, welches die Generalisierbarkeit des reziproken Modells über die Altersgruppen hinweg unterstützen würde (Guay, Marsh, & Boivin, 2003). Daher wurde trotz des wachsenden Supports für das reziproke Modell weiterhin über die Richtung der Beziehung zwischen akademischem Selbstkonzept und Schulleistung diskutiert (Wigfield & Wagner, 2007).

Einen substantiellen Beitrag zur Debatte über die kausale Anordnung von Selbstkonzept und Schulleistung bietet die Metaanalyse zu Längsschnittstudien von Huang (2011). Bis dahin waren einzig vier Metanalysen zu Studien mit Querschnittsdesign erschienen. In einer ersten Metaanalyse von Hansford und Hattie (1982) wurde systematisiert, in welcher Beziehung unterschiedliche Selbstvorstellungen zu Leistungsmessungen stehen. Hierzu wurden neben dem Selbstkonzept auch Konstrukte wie das Selbstwertgefühl oder die Selbstakzeptanz einbezogen. Aus den beiden häufig zitierten Metaanalysen von Valentine (Valentine, 2001; Valentine, DuBois, & Cooper, 2004) können nur bedingt Schlussfolgerungen abgeleitet werden, da erstens überlappende Stichproben einbezogen wurden. Zudem wurden die Effektgrößen höchst wahrscheinlich unterschätzt, da standardisierte Regressionskoeffizienten als Maß für die Effektgrößen verwendet wurden. Folglich musste angenommen werden, dass die Varianz der Schulleistung – welche im Regressionsmodell mit dem Selbstkonzept und anderen erklärenden Variablen geteilt wurde – auch auf andere Prädiktoren als dem Selbstkonzept zurückzuführen war. Die zusätzlichen Analysen zeigten denn auch, dass die Effekte statistisch abhängig von der Anzahl der Prädiktoren waren (siehe hierzu Kritik von Huang, 2011). Die Metaanalyse von Möller und Mitarbeitenden (2009) schließlich überprüfte explizit die Generalisierbarkeit des *internal/external frame of reference* beziehungsweise I/E-Modells (u. a. Köller, Klemmert, Möller, & Baumert, 1999; Marsh & Köller, 2003; Skaalvik & Skaalvik, 2002), welches auf der konzeptionellen Unterscheidung zwischen mathematischem und verbalem akademischem Selbstkonzept und damit auf dem Marsh/Shavelson-Modell basiert. Die Resultate verwiesen auf starke Zusammenhänge zwischen verbalem Selbstkonzept und verbaler Leistung sowie mathematischem Selbstkonzept und mathematischer Leistung. Die Metaanalyse unterstützt damit das I/E-Modell, wonach interne Vergleiche zwischen mathematischen und verbalen Schulleistungen einen negativen Effekt auf das akademische Selbstkonzept des jeweils anderen Fachbereiches haben (*internal frame of reference*) und soziale Vergleiche zu positiven Effekten auf das entsprechende fachspezifische Selbstkonzept haben (*external frame of reference*).

In der oben angekündigten Metaanalyse von Huang (2011) wurden hingegen 32 Längsschnittstudien mit 39 unabhängigen Stichproben einbezogen. Das harmonische Mittel über die verschiedenen Stichprobengrößen hinweg betrug 136 Testpersonen; 19 Studien erhoben Daten an 2 Messzeitpunkten, 8 Studien an 3 Messzeitpunkten

(weitere deskriptive Angaben zu den einbezogenen Studien siehe Huang, 2011, S. 513). Die auf Stichprobengröße sowie Unreliabilität korrigierten Korrelationen (nach Hunter & Schmidt, 2004) zeigten im Mittel einen Zusammenhang von .20 bis .27 zwischen vorherigem Selbstkonzept zu nachfolgender Schulleistung sowie im Mittel einen Zusammenhang von .19 bis .25 zwischen vorangegangener Leistung zu nachfolgendem Selbstkonzept. Ebenfalls moderat bis groß (basierend auf Hattie, 2009) war die durchschnittliche longitudinale Korrelation zwischen Selbstkonzept und Schulleistung. Die Ergebnisse zeigen demnach, dass zwischen dem akademischem Selbstkonzept und der Schulleistung eine positive Wechselbeziehung besteht.

Welche praktische Bedeutung hat nun dieses moderate bis große Ausmaß der durchschnittlichen Effektgröße? Zur Visualisierung eignet sich der *Binomial Effect Size Display* (BESD) nach Rosenthal und Rubin (Lipsey & Wilson, 2001; Rosenthal & Rubin, 1982): Wird das akademische Selbstkonzept anhand des Medians unterteilt in „tief versus hoch“ sowie die Schulleistung in „Erfolg versus Misserfolg“ bedeutet eine Korrelation von .20 eine 60 Prozent Erfolgsquote für jene mit einem hohen Selbstkonzept und eine 40 Prozent Erfolgsquote für jene mit einem tiefen Selbstkonzept – oder umgekehrt eine 60 Prozent Erhöhungsquote des Selbstkonzepts für die leistungsstärkere Hälfte und 40 Prozent Erhöhungsquote des Selbstkonzepts für die leistungsschwächere Hälfte.

Eine weitere zentrale Absicht der Metaanalyse von Huang (2011) war die Identifizierung von generellen Moderatoren respektive Interaktionseffekten. Dabei stellte sich der Faktor „Globalität versus Spezifität“ als einziger signifikanter Moderator in den Beziehungen zwischen (vorherigem) Selbstkonzept und (nachfolgender) Schulleistung sowie zwischen (vorangegangener) Leistung und (nachfolgendem) Selbstkonzept heraus. Konkret bedeutet dies, dass in Studien, bei denen das globale Selbstkonzept erfasst wurde, relativ schwache Korrelationen ersichtlich wurden, im Vergleich zu Studien, in denen das akademische oder fachspezifische Selbstkonzept untersucht wurde (siehe auch Marsh & Craven, 2006).

Diese Ergebnisse dürfen indessen nicht darüber hinwegtäuschen, dass ein nicht unerheblicher Anteil der Varianz von Selbstkonzept und Schulleistung weiterhin unerklärt bleibt. So konnten in den 32 Längsschnittstudien nur zwischen vier bis sieben Prozent an Varianz der Schulleistung durch das Selbstkonzept et vice versa erklärt werden. Deshalb sollten in zukünftigen Untersuchungen weitere relevante Merkmale zur Vorhersage von Schulleistung einbezogen werden (Huang, 2011).

2.3 Selbstkonzept und soziale Vergleiche aus sonderpädagogischer Perspektive

Dem Selbstkonzept kommt in der sonderpädagogisch ausgerichteten Forschung seit nunmehr über vier Jahrzehnten große Beachtung zu. Einen Eckstein diesbezüglich lieferte Ende der 1980er-Jahren James W. Chapman (1988a) in einem Forschungsüberblick zu Untersuchungen zum Selbstkonzept von Schülerinnen und Schülern mit einer sogenannten „Lernbehinderung“. Einbezogen wurden über 21 Studien zum generellen Selbstkonzept und 20 Studien zum akademischen Selbstkonzept. Dabei stand Chapman vor der grundlegenden Problematik der Operationalisierung des Begriffs *learning disabilities*: In die Analyse einbezogen wurden schließlich Studien mit Stichproben, bei denen die Klassifizierung *learning disabilities* oder ein ähnliches diagnostisches Label (z. B. *retarded readers* oder *mildly handicapped*) bereits von Seiten der Schulbehörden vorgenommen worden war sowie jene, die von Seiten der Forschenden anhand einer vergleichbaren Operationalisierung als *learning disabled* identifiziert worden waren. Von den Schulbehörden als *learning disabled* bezeichnet wurden im Übrigen Schülerinnen und Schüler, bei denen eine signifikante Diskrepanz zwischen gemessenen Leistungen und Fähigkeiten bestand und deren Lernschwierigkeiten nicht einer mentalen Retardation oder gewichtigen sensorischen, physischen, emotionalen oder umweltbezogenen Faktoren zuzuschreiben waren. Zusammenfassend kam Chapman zu dem Ergebnis, dass Schülerinnen und Schüler mit *learning disabilities* generell ein tieferes Selbstkonzept als andere Schülerinnen und Schüler aufwiesen. Beim akademischen Selbstkonzept waren dabei gegenüber dem generellen Selbstkonzept eindeutig größere Unterschiede auszumachen.

Bezüglich Schulform wurden von Chapman erstaunlicherweise keine Effekte festgestellt. Allerdings verzeichneten „lernbehinderte“ Schülerinnen und Schüler in Regelklassen, die keine sonderpädagogische Unterstützung erhielten, ein tieferes Selbstkonzept als solche, die sonderpädagogischen Förderunterricht erhielten. Chapman (1988b) warnte daher im Zuge der „pädagogischen Normalisierung“ in Neuseeland davor, Schülerinnen und Schüler mit „besonderen Bedürfnissen“ ohne zusätzliche Unterstützung in Regelschulen einzugliedern – mit der Begründung, *mainstreaming* nicht zum *maindumping* verkommen zu lassen.

2.3.1 Pioniere der Selbstkonzeptforschung und bezugsgruppentheoretischer Ansätze

Die empirische Befundlage zu Effekten der Beschulungsform war in den Anfängen der Selbstkonzeptforschung tatsächlich inkonsistent. Bei Meyerowitz (1962) beispielsweise war bei Schülerinnen und Schülern, welche als *educable mentally handicapped* galten, nach ihrem ersten Schuljahr in einer Sonderklasse ein (wohl-gemerkt) tieferes Selbstkonzept zu erkennen als bei vergleichbaren Schülerinnen und Schülern in Regelklassen (ähnliche Befunde u. a. auch bei Piers & Harris, 1964).

Andere Studien wiederum rapportierten Befunde mit umgekehrten Vorzeichen. Dazu gehören etwa Rheinberg und Enstrup (1977), welche im deutschsprachigen Raum zu den ersten zählten, die sich der Fragen nach der Wirkung der Schulform auf das Selbstkonzept annahmen. Mit den Ergebnissen aus ihrer Querschnittsstudie lieferten sie Belege für die Hypothese von Wulf-Uwe Meyer (1973), wonach das „Selbstkonzept der Begabung“ entscheidend von sozialen Vergleichen innerhalb der Bezugsgruppe bestimmt werde. Schülerinnen und Schüler aus Sonderschulen wiesen dabei ein höheres Selbstkonzept als Regelklassenschülerinnen und -schüler mit einem vergleichbaren Intelligenzquotienten auf. Zudem ergaben die Auswertungen eine Interaktion zwischen Schulform und Klassenstufe: Die auf Klassenstufe 4 bis 7 starken schulformspezifischen Effekte wurden auf Klassenstufen 8 und 9 aufgehoben (Rheinberg & Enstrup, 1977). Die Hauptbefunde fanden im selben Jahr Unterstützung mit der Längsschnittstudie von Krug und Peters. Bei den 40 in die Stichprobe einbezogenen Schülerinnen und Schülern, welche zu Beginn der dritten oder vierten Klasse in eine Sonderschule überwiesen worden waren, stieg im Verlaufe des Schuljahres das Selbstkonzept stark an, während es bei den in der Regelklasse verbliebenen Schülerinnen und Schülern mit vergleichbarem Intelligenzquotienten stagnierte (Krug & Peters, 1977).

Erstaunliche und deswegen vielfach aufgeführte Ergebnisse lieferte das Experiment von Strang, Smith und Rogers (1978), bei dem ein Teil der 50 in die Untersuchung einbezogenen Sonderklassenschülerinnen und -schüler (bezeichnet als *academically handicapped children*) halbtags in eine Regelklasse integriert wurden. Das Selbstkonzept der Schülerinnen und Schüler der Experimentalgruppe stieg im Vergleich zum Selbstkonzept der Kontrollgruppe, mit den in Sonderklassen verbliebenen Schülerinnen und Schülern, über drei Messzeitpunkte signifikant an. Als mögliche Begründung führte die Autorengruppe auf, dass die Schülerinnen und Schüler für die „akademisch-relevanten sozialen Vergleiche“ vermutlich jene Schülerinnen und Schüler herangezogen, welche ihnen am ähnlichsten seien – womit die Gruppe der in der Sonderklasse verbliebenen „akademisch behinderten Kinder“ gemeint waren. Strang, Smith und Rogers untersuchten überdies zwei

Jahre später anhand eines weiteren Experiments, wie sich das Selbstkonzept bei Veränderung der Referenzgruppe verhielt: Die inzwischen alle halbtags in die Regelklasse integrierten ehemaligen Sonderschülerinnen und Sonderschüler schätzten ihre Fähigkeiten während der Situation in der Sonderklasse höher ein als während der Regelklassensituation (ebd.). Aus diesem Befund schlossen Schwarzer, Lange und Jerusalem (1982), dass beim aus ihrer Sicht „unrealistischen“ Vergleichen mit einer leistungsbesseren Gruppe (der Regelklasse) eine „relative Deprivation“ erfahren werde, und beim Vergleich mit einer leistungsschwächeren Gruppe (der Sonderklasse) eine „relative Gratifikation“.

Mit *relativer Deprivation* und *relativer Gratifikation* sind jene bezugsgruppentheoretischen Begriffe angesprochen, welche vor allem in den 1960er- und 1970er-Jahren in der relevanten Fachliteratur gehäuft anzutreffen waren. Ursprünglich von Samuel A. Stouffer und Mitarbeitenden (1949) zur Erklärung unterschiedlicher Einstellungen von Soldaten in der Studie „*The American soldiers*“ verwendet, fanden diese Konstrukte später Einzug in den pädagogischen Forschungsbereich. So bezogen sich etwa zwei Untersuchungen im College-Bereich auf die Theorie der relativen Deprivation, und zwar die in Abschnitt 2.1 bereits aufgeführte Studie von Davis (1966) – nota bene erschienen unter dem Titel „*The campus as a frog pond*“ – sowie jene von Werts und Watley (1969) – mit dem Titel „*A student's dilemma: big fish – little pond or little fish – big pond*“.

Der Soziologe Robert K. Merton nahm sich die empirischen Untersuchungen der Autorengruppe um Stouffer zum Anlass, in Zusammenarbeit mit Rossi (ehemals Kitt) die einzelnen Studien eingehend zu analysieren, mit dem Ziel theoretische Implikationen abzuleiten und dadurch die Theorie des Referenzgruppenverhaltens zu erweitern (Merton & Rossi, 1968). Ihre Umgestaltung des Theorieansatzes zu einer umfassenderen Referenzgruppentheorie konnte sich jedoch im pädagogischen Forschungsfeld nicht etablieren und wurde in seiner Umsetzung als zu unscharf kritisiert. Unklar blieb unter anderem, ob die relative Deprivation als Folge eines sozialen Vergleichs oder als intervenierender Faktor zu fassen war, und ob es sich dabei um eine Kognition oder eine Emotion handelte (für eine kritische Auseinandersetzung siehe J. W. L. Wagner, 1999).

Zu den Ausnahmen, welche sich der bildungssoziologischen Referenzgruppentheorie von Merton und Rossi bedienten, zählte Richer (1976). Bei seiner Übertragung der Theorie auf den Kontext von Leistungsgruppierung ging er von einem Einfluss der schulischen Referenzgruppe auf kognitiver Ebene aus. Mit Verweis auf Merton (1968a) und Kelley (1955) bezog er unter anderem das Konzept *salience* respektive Salienz mit ein, welches das wahrnehmungsbezogene Herausragen eines Merkmals oder Objekts aus seiner Umgebung beinhaltet. Die beiden Komponenten *Sichtbarkeit* und *Bedeutsamkeit* einer Person wurden von Richer als sowohl notwendige als

auch ausreichende Bedingungen für die Herstellung von Referenzgruppenbezügen angegeben. Das Konzept Salienz nahm nebenbei bemerkt bereits in Festingers Theorie sozialer Vergleichsprozesse einen wichtigen Stellenwert ein.

In einem weiteren theoretischen Themenkomplex bezog sich Richer auf die Art der Beeinflussung durch die Bezugsgruppe. Dabei unterschied er – in Anlehnung an die erwähnten Protagonisten – zwischen *komparativen* und *positiv normativen* (via Identifikation mit den Gruppennormen) sowie *negativ normativen* (via Ablehnung der Normen) Referenzgruppeneffekten. Die postulierten Konsequenzen von Leistungsgruppierung für Schülerinnen und Schüler mit schwachen Schulleistungen fasste er folgendermaßen zusammen:

Advocates of homogenous grouping argue, though not in these words, that the low-ability student in a heterogeneous class takes high-ability students as a comparative reference group. The results are assumed to be feelings of inferiority and inadequacy. A homogeneously grouped low-ability student fares better because such feelings do not occur. Those opposing homogeneous grouping, however, hold that a low-ability student in a heterogeneous class emulates and learns from his high-ability peers. (Richer, 1976, p. 69)

Rückblickend mag erstaunen, dass Richers recht differenzierte Hypothesen zur Erklärung von Referenzgruppenprozessen in (sonder-)pädagogischen Themenbereichen selten rezipiert wurden. Stattdessen setzten sich Ansätze in Orientierung an Festinger im vorwiegend sozialpsychologisch besetzten Forschungsfeld zu sozialen Vergleichsprozessen durch – zumal sich diese für die Erklärung des Selbstkonzepts vortrefflich eigneten.

2.3.2 Soziale Vergleiche im Fokus der sonderpädagogisch ausgerichteten Forschung

Der aufmerksamen Leserin respektive dem aufmerksamen Leser wird vermutlich nicht entgangen sein, dass die im vorangehenden Abschnitt zitierten Studien zum Selbstkonzept in der Blütezeit der *Dust Bowls* der Selbstkonzeptforschung (vgl. Abschnitt 2.2.1) durchgeführt worden waren. Forschende bemühten sich daher in den darauffolgenden Jahren, methodische Mängel von Studien der Selbstkonzept-Pioniere aufzudecken und das Vorgehen in ihren eigenen Studien zu optimieren (z. B. E. J. Cooley & Ayres, 1988; Rogers & Saklofske, 1985). So kritisierten beispielsweise Renick und Harter (1989) unter anderem die Studien von Strang und Mitarbeitern (1978): Zum einen werde mit dem verwendeten Instrument (Piers-Harris' *children's self-concept scale*) einzig ein globales Selbstwertgefühl gemessen (vgl. auch Kritik

von Chapman, 1988a). Zum anderen sei nicht klar, mit welcher Gruppe sich die Kinder vergleichen würden. Um diese Frage zu klären, führten Reineck und Harter eine Untersuchung mit 86 Schülerinnen und Schülern der dritten bis achten Klassenstufe durch, welche während einer Lektion pro Schultag in einer kleinen Gruppe in einem separaten Raum (*learning disabilities resource room*) speziell unterrichtet wurden. Die Schülerinnen und Schüler wurde einzeln befragt und jeweils darum gebeten anzugeben, mit welcher der beiden Gruppen sie sich vergleichen würden. Die Auswertungen ergaben, dass die Schülerinnen und Schüler während der Regelklassensituation ihre schulischen Fähigkeiten tiefer einschätzten als während dem Unterricht im separativen Raum. Als „interessant“ beurteilten Reineck und Harter den Befund, wonach sich 84 Prozent der zusätzlich geförderten Schülerinnen und Schüler spontan mit ihren Regelklassenkameradinnen und -kameraden verglichen. Interessant daher, weil dies zum einen im Gegensatz zu den Befunden von Silon und Harter (1985) stand, und zum anderen, weil dies ihrer Ansicht nach einer der zentralen und häufig aufgeführten Hypothesen aus Festingers sozialer Vergleichstheorie widersprach. Aus der Anzahl im Umfeld vorhandenen Personen werden nämlich laut Festinger (1954) jene für den Vergleich gewählt, welche ähnliche Fähigkeiten haben. Eine ihrer Schlussfolgerungen für weitere Untersuchungen zu sozialen Vergleichen lautete daher, es sei wichtiger, die wahrgenommene Ähnlichkeit zu berücksichtigen als lediglich von einer möglichen Referenzgruppe auszugehen.

Der Gedanke, dass das Selbstkonzept in hohem Maße abhängig von sozialen Vergleichen innerhalb der unmittelbaren – oder eben sichtbaren und bedeutsamen Referenzgruppe – ist, und nicht von den Personen, die nach objektiven Kriterien leistungsmäßig am ähnlichsten sind, begann sich im Verlaufe der 1980er-Jahre nach und nach durchzusetzen (Hinweise dafür z. B. bei Bachman & O'Malley, 1986; J. M. Coleman, 1983; Silverman & Zigmond, 1983). Den größten Beitrag diesbezüglich ist jedoch wohl unangefochten Marsh zuzuschreiben (u. a. Marsh, 1987; Marsh & Parker, 1984).

Auch heute noch erhält Marsh innerhalb der Forschungsarbeiten zu sozialen Vergleichen am meisten Beachtung. Überdies wird er als produktivster Autor der pädagogischen Psychologie aufgeführt (z. B. Marsh, Craven, & McInerney, 2008). Sein *big-fish-little-pond effect* erklärt auf einleuchtende Art und Weise die Relativität des akademischen Selbstkonzepts in Bezug zur jeweiligen schulischen Referenzgruppe. Die Beliebtheit des Fischteich-Effekts gründet zudem auf seiner guten empirischen Absicherung. Vielfach repliziert gilt er mittlerweile als länderübergreifend generalisierbar (Marsh & Hau, 2003; Seaton, Craven, & Marsh, 2008).

Um Effekte der gesonderten Beschulung zu erklären, erweiterte Marsh gemeinsam mit Johnston (1993) die theoretische Basis seines referenzgruppentheoretischen Modells. Eine empirische Überprüfung des erweiterten Modells erfolgte rund zehn

Jahre später in Zusammenarbeit mit Tracey und Craven (2006; 2003). Die Resultate der mit 211 Schülerinnen und Schülern der zweiten bis sechsten Grundschulstufe durchgeführten Studie ließen erkennen, dass der BFLPE ebenfalls für Schülerinnen und Schüler mit *mild intellectual disability* gilt, und unterstützten damit die Vorhersage, wonach sich das akademische Selbstkonzept bei einer Sonderbeschulung im Vergleich zu einer Regelbeschulung erhöht.

Diese Befunde vermögen hierzulande wohl kaum allzu viel Aufsehen erregen. Eher ungewöhnlich ist hingegen das von der Autorenschaft eingebrachte Argument, wonach das höhere akademische Selbstkonzept in Sonderschulen als Widerspruch zur *labelling theory* respektive zum Etikettierungs-Ansatz stehen würde. Damit zielen sie mitten in die bereits seit Mitte des letzten Jahrtausends sowohl in Politik, Praxis als auch Forschung zuweilen hitzig geführte Integrations-Separations-Debatte.

2.3.3 Selbstkonzept als Argument in der Integrations-Separations-Debatte

Results support the big fish little pond effect's prediction that special class placement enhances the academic self-concepts of students with mild intellectual disabilities. At the same time these findings contradict labelling theory – the basis of much current special education philosophy. (Tracey et al., 2003, p. 224)

Tatsächlich haben Proponenten der Integrationsbewegung seit den 1970er-Jahren den Etikettierungsansatz und in diesem Zusammenhang das Stigma-Konzept nach Erving Goffman (1963) des Öfteren in ihre Argumentation für das Normalisierungsprinzip und gegen die Ausdifferenzierung des Sonderschulwesens herangezogen (vgl. u. a. Haeberlin, 2005). Aufsehererregend war im englischen Sprachraum der Artikel von Lloyd Dunn (1968), der die Sonderklassen für die als *mentally retarded* diagnostizierten Schülerinnen und Schüler nicht nur als ineffizient sondern auch als unhaltbar beurteilte. Unhaltbar deshalb, weil das erklärte Ziel des Diagnoseprozesses zwar eine ganzheitliche Betrachtung eines Kindes beabsichtige, es jedoch lediglich darum gehe, es mit einem Label wie *mentally retarded*, *emotionally disturbed*, *minimally brain injured* oder – je nach Fachrichtung der diagnostizierenden Personen – mit einem anderen Begriff zu versehen, welches eine Sonderklassenplatzierung rechtfertigen würde. Für nicht wenige Kinder aus ethnisch oder ökonomisch benachteiligten Familien würde dieses einvernehmliche Täuschungsbündnis („*entente of mutual delusion*“) zwischen Regel- und Sonderpädagogik einem Abschieben gleichkommen (ebd., S. 20). Folgen dieser Etikettierungen seien eine verminderte Leistungserfolgserwartung auf Seiten der Lehrpersonen (vgl. Rosenthal & Jacobson, 1966) sowie auf Seiten des betroffenen

Kindes möglicherweise Gefühle der Minderwertigkeit und Probleme der Akzeptanz. Von der pädagogisch-psychologischen Selbstkonzeptforschung aufgegriffen wurde Dunns Aussage, Sonderbeschulung habe wahrscheinlich einen ernstzunehmenden hinderlichen Effekt auf das Selbstbild im allgemeinen Sinne (L. M. Dunn, 1968). In Anlehnung an solche und ähnliche Aussagen wurde das höhere Selbstkonzept in Sonderklassen mitunter als Beweisgrund gegen die Stigmatisierungs-Theorie der Integrationsbefürworter verwendet (u. a. Crabtree, 2003; Rheinberg & Enstrup, 1977). Dabei übersahen die Autoren geflissentlich, dass *erstens* der Begriff Stigma immer in einer „Sprache von sozialen Beziehungen“ zu verwenden ist und nicht per se ein diskreditiertes Merkmal darstellt (Goffman, 1963, S. 3), *zweitens* eine solche Stigmatisierung gerade durch die Sonderklassenzuweisung erfolgt (z. B. Thimm, 1975), und *drittens* weder das generelle noch ein spezifisches Selbstkonzept kompatibel mit der Stigma-Theorie sind (für eine Auseinandersetzung zum Stigma-Konzept siehe Link & Phelan, 2001).

Der Ursprung für die Unstimmigkeiten liegt vermutlich darin, dass Vertretende des Etikettierungsansatzes und Forschende zum Selbstkonzept unterschiedlichen Fachbereichen beziehungsweise wissenschaftlichen „Denkgruppen“ angehören, und dadurch eine Verständigung erschwert ist oder gar verunmöglicht wird (vgl. Fleck, 2008).

Stigmatisierung ist auch heutzutage noch eines der wichtigen Argumente in der inzwischen vornehmlich bildungspolitisch geführten Diskussion um Separation versus Integration oder Inklusion. (Auf die Debatte um die beiden im pädagogischen Bereich meist hierarchisch verwendeten Begriffe Integration und Inklusion wird in dieser Arbeit nicht eingegangen.) Die jüngere Integrationsforschung beschäftigt sich jedoch vornehmlich mit dem Selbstkonzept im Allgemeinen und dem akademischen Selbstkonzept im Besonderen. In einem Forschungsüberblick zu Wirkungsweisen von integrativen im Vergleich zu separativen Schulformen zeigt sich hinsichtlich dem akademischen Selbstkonzept eine konsistente Befundlage, nämlich „dass integrierte lernbehinderte Schülerinnen und Schüler im Vergleich zu lernbehinderten Sonderschülern ein niedrigeres Selbstkonzept hinsichtlich ihrer *schulischen* Leistungsfähigkeiten haben“ (Bless & Mohr, 2007, 381, Hervorh. i. Orig.; vgl. auch Ruijs & Peetsma, 2009). Bezüglich des allgemeinen Selbstkonzeptes ergibt sich ein eher widersprüchliches Bild, was sich unter anderem mit der inkonsistenten Begriffsverwendung erklären lässt. Des Weiteren ergaben sich bezüglich dem allgemeinen Selbstkonzept *innerhalb* der „Gruppe“ der Schülerinnen und Schüler mit „Lernbehinderung“ ähnlich große Unterschiede wie *zwischen* Schülerinnen und Schülern mit und ohne „Lernbehinderung“ (Crabtree, 2003). Kinder oder Jugendliche scheinen eine Vorstellung von sich selbst generell betrachtet demnach nicht aufgrund ihrer schulischen Leistungen aufzubauen – was der Differenzierung

zwischen verschiedenen Selbstkonzeptdomänen entspricht. Dass jedoch Schülerinnen und Schüler, welche schwache Schulleistungen aufweisen, ihre schulischen Fähigkeiten tiefer als der Durchschnitt einschätzen, kann als logisch – oder „realistisch“ – betrachtet werden (z. B. Zeleke, 2004).

Das Stichwort „realistisch“ ist höchstwahrscheinlich in nahezu jeder Diskussion um Integration versus Separation anzutreffen, die in etwa so verlaufen könnte: Das schulische Selbstkonzept sei eben tiefer, dafür aber realistischer, sagen die Integrationsbefürworter, und schließlich müssten die Jugendlichen später in der realen Berufswelt bestehen können. Ein tiefes schulisches Selbstkonzept sei aber abträglich für die Entwicklung der Jugendlichen, erklären die Integrationsgegner. Worauf die Kontrahenten entgegnen, dass dem ungünstigen Effekt auf das akademische Selbstkonzept mit entsprechenden pädagogischen Praktiken begegnet werden könne (z. B. Strein, 1993).

Forschungsarbeiten zum akademischen Selbstkonzept zielen daher häufig darauf ab pädagogische Interventionen zur Erhöhung des Selbstkonzepts zu untersuchen. Elbaum und Vaughn unterstreichen in ihren Metaanalysen (2001, 2003) die Wichtigkeit entsprechender Interventionen primär für Schülerinnen und Schüler mit schwachen Schulleistungen *und* einer tiefen Einschätzung ihrer schulischen Fähigkeiten, da diese mit Abstand am meisten von einer Unterstützung profitieren würden (2003) – eine Feststellung die trivial erscheint. In einer kritischen Stellungnahme zu den Ergebnissen aus den beiden Metaanalysen fügt Bracken (2009) an, dass die tiefen Effektstärken die Vermutung nahelegen würden, es handle sich um eine artifizielle Verbesserung des akademischen Selbstkonzepts. Seine Empfehlung, die direkteste Methode für dessen Erhöhung sei es wohl, die Schülerinnen und Schüler mit schwachen Schulleistungen darin zu unterstützen, bessere Leistungen zu erreichen, wirkt da schon fast etwas lapidar.

Anbetracht dessen, dass direkte neben indirekten Leistungsrückmeldungen als zentrale Informationsquellen für die Selbsteinschätzung der schulischen Fähigkeiten dienen (u. a. Filipp, 2006), und dementsprechend Leistungsbeurteilungen der Lehrpersonen das akademische Selbstkonzept von Schülerinnen und Schülern bedeutsam beeinflussen (u. a. Bear, Minke, Griffin, & Deemer, 1998; Wilbert & Gerdes, 2009), eröffnet Brackens Vorschlag eine neue Dimension – die im nun folgenden Abschnitt 2.4 im Zusammenhang mit schulischen Selektionspraktiken und Leistungsvergleichsprozessen thematisiert wird.

2.4 Referenzgruppeneffekte im Kontext von Selektion und ihre Folgen

Untersuchungen zu sozialen Vergleichsprozessen befassen sich größtenteils mit dem akademischen Selbstkonzept von Schülerinnen und Schülern mit schwachen Schulleistungen in den folgenden zwei Bereichen, die beide mit Selektionsprozessen in Verbindung stehen: entweder im Kontext der vorangehend thematisierten Beschulungsform von Schülerinnen und Schülern mit „Lernbehinderung“ oder im Kontext von Leistungsgruppierung auf der Sekundarstufe I (Byrne, 1996).

Eine der ersten Studien, die sich – wie auch diese Arbeit – mit der Selbstkonzeptentwicklung im Zusammenhang mit der zweit genannten Selektionsschwelle im Schweizerischen Bildungssystem befasst hatte, war jene von Buff (1991). Konkret wurde bei Buff die Selbstkonzeptentwicklung von Schülerinnen und Schülern im Kanton Bern über vier Messzeitpunkte vor und nach dem Schulübertritt untersucht. Dieser Übergang fand damals bereits nach dem vierten Primarschuljahr statt.

Hinsichtlich der Selbstkonzeptentwicklung scheinen die „negativ Selegierten“ vom Wechsel in die Realschule zu profitieren. Zumindest zu Beginn der Sekundarstufe I stellen die Realschüler im Schnitt noch keine von zu wenig Selbstvertrauen und tiefem Selbstwertgefühl geplagte Schülerpopulation dar.

Diesen Sachverhalt als Argument für separierende Maßnahmen im Schulwesen anzuführen ist jedoch insofern fragwürdig, als es sich bei den bezugsgruppentheoretisch zu interpretierenden günstigen Entwicklungen unter Umständen um ein kurzfristiges Phänomen handeln könnte, das primär auf die unteren Klassenstufen der Sekundarstufe I beschränkt ist. ... Dass sich die befürchteten negativen Entwicklungen allenfalls doch noch einstellen, wenn die Realschüler zu merken beginnen, welche Zukunftsoptionen ihnen verschlossen sind, ist durchaus möglich. (Buff, 1991, S. 112)

Die Ergebnisse von Buff verweisen zum einen auf komparative Referenzgruppeneffekte, welche durch die vorgängige Selektion mitbestimmt werden. Abschnitt 2.4.1 beleuchtet Referenzgruppeneffekte vor dem Hintergrund eines selektiven Bildungssystems und nimmt dabei auch die damit verbundene Leistungsbeurteilung in den Fokus. Die institutionalisierten Zuweisungsprozesse haben sowohl kurz- als auch langfristige Folgen, die vor allem für Schülerinnen und Schüler mit schwachen Schulleistungen ins Gewicht fallen und daher – ähnlich einem Exkurs – in Abschnitt 2.4.2 behandelt werden. Abschnitt 2.4.3 greift schließlich die von Buff geäußerte Befürchtung auf und beabsichtigt damit gleichzeitig eine Erläuterung der bereits in der Einleitung aufgeworfenen Frage nach komparativen Bezugsgruppeneffekten auf Bildungs- und Berufsaspirationen.

2.4.1 Leistungsbeurteilung, Leistungsgruppierung und Referenzgruppeneffekte

Die folgenreichste Form von Leistungsbeurteilung ist jene, die üblicherweise in Form von Noten ausgestellt wird. Festgelegte Schwellenwerte innerhalb des Notenspektrums dienen dazu, den Übertritt in die nächste Klasse oder Stufe zu gewähren oder die Zuweisung in einen Schultyp zu bestimmen. Doch was bereits im 19. Jahrhundert empirisch unterlegt kritisiert (Edgeworth, 1888; zitiert nach Kronig, 2007, S. 11) und von Ingenkamp (1969, 1971a) als einer der ersten deutlich und empirisch unterlegt bezweifelt wurde, ist von Kronig (2007) eindrücklich belegt worden: Noten verlieren ihre Aussagekraft außerhalb der einzelnen Schulklasse. Während man beispielsweise in einem Schweizer Kanton zu den besten 60 Prozent der Schülerschaft gehören muss, um am Ende der Primarschulzeit in einen Schultyp mit erweiterten Ansprüchen versetzt zu werden, genügt es in einem anderen Kanton, zu den besten 90 Prozent zu gehören. Konkret kann dies heißen, dass in acht von zehn Fällen ein Schüler oder eine Schülerin mit vergleichbaren Leistungen im jeweils anderen Schultyp vorzufinden ist. Überdies ist auch die Zuweisung in eine Sonderklasse eindeutig standortabhängig: In einem Kanton ist jeder zweihundertste davon betroffen, in einem anderen jeder fünfundzwanzigste (ebd.).

Je nach Leistungsniveau einer Schulklasse können Schülerinnen und Schüler für dieselbe Leistung völlig unterschiedliche Noten erhalten (Kronig, 2008). Die jeweilige Klasse dient sozusagen als Maßstab für die Leistungsbeurteilung, die wiederum über die Zuteilung zu einem Schultyp bestimmen soll. Anbetracht dieser „unglaublich gewordenen Selektion“ (Kronig, 2007, S. 226), mutet es schon fast als Anmaßung an, wenn im Zusammenhang mit schwachen Schulleistungen von pädagogischen Interventionen zur Steigerung des akademischen Selbstkonzepts gesprochen wird. Vielmehr scheint die schulische Selektion als strukturelle Maßnahme einen intervenierenden Faktor bei der Formierung des akademischen Selbstkonzepts darzustellen.

Ein vielfach verwendetes Modell zur Erklärung von Wirkungsweisen der Leistungsgruppierung ist bekanntlich der *big-fish-little-pond effect*. Bisher wurden hierzu fast ausschließlich Studien aus dem englischen Sprachraum zitiert. Daher stellt sich die Frage, ob dieser Referenzgruppeneffekt ebenfalls in Bildungssystemen mit früh angesetzter Selektion – also in Deutschland und in der Schweiz – gültig ist.

Ein außergewöhnliches quasi-experimentelles Untersuchungsdesign zur Überprüfung von Effekten der Leistungsgruppierung bot sich mit der Deutschen Wiedervereinigung. Bis 1991 existierten in den beiden Ländern zwei unterschiedliche Bildungssysteme: Während im Ostdeutschen System über die gesamte obligatorische Schulzeit in heterogen zusammengesetzten Klassen unterrichtet wurde, erfolgte im

Westdeutschen System ab dem fünften Schuljahr eine Unterteilung in verschiedene Schultypen. Im Zuge der Wiedervereinigung wurde dieses institutionelle Selektionsprinzip für alle Schulen in Deutschland übernommen.

Für die entsprechenden Analysen verwendeten Marsh, Köller und Baumert (Marsh, Köller, & Baumert, 2001) eine Teilstichprobe der Längsschnittstudie „*Bildungsverläufe und psychosoziale Entwicklung im Jugendalter*“ (BIJU). Es wurden 2778 Schülerinnen und Schüler der siebten Jahrgangsstufe aus 161 Klassen einbezogen, von denen sich demzufolge jene aus Westdeutschland seit (mindestens) zwei Jahren im Bildungssystem mit Leistungsgruppierung befanden. Dabei wurde über die drei Messzeitpunkte hinweg der negative Kontrasteffekt des Leistungsdurchschnitts der Schulklasse auf das individuelle akademische Selbstkonzept deutlich ersichtlich. Die Befunde replizierten damit nicht nur die bisherigen Forschungsbefunde zum *big-fish-little-pond effect* aus dem englischen Sprachraum (für einen Überblick siehe Marsh & Craven, 2002), gleichzeitig zeigten sie, dass der Fischteich-Effekt in Schulen des ehemaligen Ostdeutschlands zu Beginn der Umstellung auf das selektive Bildungssystem signifikant weniger stark ausgeprägt war als in ehemals Westdeutschen Schulen. In der Mitte des siebten Schuljahres war die Differenz bereits geringer, und am Ende konnte kein Unterschied mehr zwischen den beiden vormals verschiedenen Schulsystemen ausgemacht werden (zu ähnlichen Befunden bezüglich Stabilisierung des Selbstkonzepts nach dem Schulübergang kommen Aust et al., 2010).

Insgesamt lassen die Resultate erkennen, wie bildungspolitische Entscheidungen auf Makroebene – namentlich die Einführung eines selektiven Schulsystems und die Institutionalisierung von differentiellen Klassenzusammensetzungen – die Bildung des akademischen Selbstkonzepts auf Mikroebene zu beeinflussen vermögen (vgl. Marsh, Köller, et al., 2001; Marsh & O'Mara, 2008, S. 105).

Auch bei weiteren Studien zum *big-fish-little-pond effect* (BFLPE) wurde jeweils im Rahmen einer Diskussion der Ergebnisse auf die Bedeutsamkeit von spezifischen Merkmalen des Bildungssystems verwiesen. Köller, Zeinz und Trautwein (2008) forderten bildungspolitische Reformen zur Erhöhung des akademischen Selbstkonzepts, um so – gemäß der reziproken Beziehung – eine Leistungsverbesserung zu erzielen. Mit Verweis auf ihre Ergebnisse, wonach der BFLPE in Abwesenheit von Zensuren respektive Schulnoten geringer war, empfahlen sie eine Umstellung von Schulnoten auf verbale Rückmeldungen. Gerade bei Schülerinnen und Schülern mit schwachen Schulleistungen sei eine Notengebung pädagogisch wenig sinnvoll (Zein, 2006; Zeinz & Köller, 2006).

In einem Punkt fand diese auf der Grundstufe durchgeführte Studie keine Übereinstimmung mit einer auf der Sekundarstufe angelegten Untersuchung von Köller und Baumert (2001): Bei Kontrolle der individuellen Ausgangsleistung in

Mathematik schnitten Schülerinnen und Schüler aus Grundstufenklassen mit geringerem Ausgangsniveau besser ab als solche aus Grundstufenklassen mit höherem Ausgangsniveau. Auf der Sekundarstufe hingegen waren die Leistungserfolge in Mathematik im Gymnasium (dem Schultyp mit den höchsten Leistungsansprüchen) bedeutend höher als in den beiden anderen Schultypen. Als eine mögliche Erklärung für die – auf den ersten Blick – divergierenden Befunde wurden die sich in Grundstufe und Sekundarstufe unterscheidenden Unterrichtskulturen aufgeführt (Köller et al., 2008). Eine naheliegende Vermutung ist außerdem, das Vorhandensein eines zum BFLPE gegenläufigen Effekts aufgrund der Schultypzugehörigkeit (vgl. auch Marsh et al., 2000). Bei Berücksichtigung des Schultyps stieg das negative Regressionsgewicht der auf Schulebene aggregierten Mathematikleistung auch tatsächlich an. Insgesamt war dieser „Schultypeffekt“ jedoch klein und nicht signifikant (Köller & Baumert, 2001). Einen positiven Assimilationseffekt im Sinne eines *reflected-glory*-Effekts (vgl. Abschnitt 2.1) als Erklärung auszuschließen wäre indessen übereilt, zumal weder der von den Schülerinnen und Schülern wahrgenommene Status der Schule noch soziale Hintergrundmerkmale berücksichtigt wurden.

Der Fischteich-Effekt wurde in Deutschland mittels des national repräsentativen Datensatzes aus der PISA-Studie 2000 (*Programme for International Student Assessment*) neben der Mathematik zudem für den Lesebereich repliziert (Lüdtke, Köller, Artelt, Stanat, & Baumert, 2002). Laut der Autorengruppe haben Schülerinnen und Schüler bei einer Zuweisung in einen Schultyp mit höheren Ansprüchen aufgrund des veränderten Bezugsrahmens mit „psychischen Kosten“ zu rechnen. Von einer Leistungsdifferenzierung sei hingegen wegen der leistungsfördernden Bedingungen in den anspruchsvolleren Schultypen nicht abzusehen. Dass jedoch eine Gruppierung nach Leistung vielfach nicht eindeutig vorgenommen werden kann, wird bei dieser Aussage nicht berücksichtigt.

Von dieser Tatsache betroffen sind nicht nur die Schülerinnen und Schüler selbst, sondern auch diejenigen Lehrpersonen, welche Empfehlungen für den Übertritt von der Primarstufe in die Sekundarstufe I abzugeben haben. So fanden Trautwein und Baeriswyl (2007) in ihrer Untersuchung zu Übertrittsempfehlungen – durchgeführt in allen deutschsprachigen sechsten Klassen des Schweizer Kantons Freiburg – Referenzgruppeneffekte, die sowohl bei der Empfehlung als auch beim tatsächlichen Übertritt substantiell über die von der Lehrperson durchgeführte Lernstandbeurteilung mediiert waren. Im konkreten Fall konnte dies zum Beispiel für einen Schüler einer insgesamt leistungsstarken Klasse eine ungünstigere Übertrittsempfehlung bedeuten, als wenn er in einer leistungsschwächeren Klasse gewesen wäre – trotz gleicher individueller Leistung in einem standardisierten Test (vgl. auch Maaz, Neumann, et al., 2008; Milek, Lüdtke, Trautwein, Maaz, & Stubbe, 2009; Tiedemann & Billmann-Mahecha, 2007). Eine stärkere Gewichtung

der Leistungstests, indem jede Lehrperson ihr eigenes Empfehlungsverhalten anhand der Leistungstests „normiert“, sowie eine höhere Plastizität bei den weiteren Schullaufbahnentscheidungen wurden als mögliche Maßnahmen vorgeschlagen, um die unerwünschten Referenzgruppeneffekte einzudämmen (Trautwein & Baeriswyl, 2007).

Die Ausrichtung an einer kriterialen Bezugsnorm zählt zu den häufigsten Empfehlungen für die Optimierung von Selektionsentscheidungen. Bereits Ingenkamp (1971b) beurteilte standardisierte Tests als das „beste gegenwärtig verfügbare Hilfsmittel“, um den Lehrpersonen „einen Vergleichsmaßstab über viele Klassen hinweg zu bieten“ (ebd., S. 163). Im Gegenzug wird aus pädagogischer Perspektive den Lehrpersonen vielfach eine individuelle Bezugsnorm nahe gelegt, um so den Schülerinnen und Schülern einen „alternativen Bezugsrahmen“ zu bieten (u. a. Köller, Schnabel, & Baumert, 2000; Lüdtke et al., 2002; Rheinberg & Krug, 2005).

Je nach Zielsetzung werden von den Lehrpersonen unterschiedliche Bezugsnormen angewandt. Eine individuelle Bezugsnorm wird hauptsächlich angewandt, wenn die Förderung einer Schülerin oder eines Schülers im Vordergrund steht (Wilbert & Gerdes, 2009). Ist das Bewertungsziel auf die Vergabe von Noten ausgerichtet, orientieren sich Lehrpersonen am klasseninternen Bezugsrahmen (Schrader & Helmke, 2002; Südkamp & Möller, 2009). Vor diesem Hintergrund stellt sich die Frage, ob eine am einzelnen Schüler beziehungsweise an der einzelnen Schülerin ausgerichtete Bezugsnorm der Lehrperson die Stärke des BFLP-Effekts einzudämmen vermag. In zwei Studien von Lüdtke und Köller (2002) zeigten sich zwar direkte positive Effekte der individuellen Bezugsnormorientierung auf das akademische Selbstkonzept in Mathematik, der negative BFLPE wurde dadurch allerdings nicht reduziert. Anscheinend können Leistungseinschätzungen von Seiten der Lehrperson, die sich an den Fortschritten des Einzelnen orientieren, Fähigkeitseinschätzungen auf Grund sozialer Vergleiche innerhalb der Schulklassse und dessen Konsequenzen nicht aushebeln.

2.4.2 Selektion und ihre Folgen

Aufgrund der aktuellen Forschungslage kann mit größtmöglicher Wahrscheinlichkeit festgehalten werden: Durch schulische Selektionsprozesse werden Schulklassen gebildet, deren Zusammensetzung sich nicht ausschließlich nach ähnlichem Schulleistungsniveau richtet.

Schon Ende der 1970er-Jahre wurde in den niedriger qualifizierenden Schultypen des Kantons Zürich ein auffallend hoher Anteil an „Fremdarbeiterknaben“ bemerkt (Häfeli, 1979). Ab Beginn der 1980er-Jahre ließ sich in der Schweizer Bildungsstatistik

eine stetig ansteigende Übervertretung ausländischer Schülerinnen und Schüler innerhalb der Klassen mit Grundansprüchen sowie in Sonder- und Kleinklassen beobachten (Lischer, 1997). Im Schuljahr 1997-1998 beispielsweise besuchte jedes zehnte Kind einer Zuwandererfamilie eine Schulklasse mit besonderem Lehrplan, von den Kindern mit Schweizer Nationalität etwa jedes vierzigste. Und dieser Umstand lässt sich nicht einfach mit schwächeren schulischen Leistungen in der Unterrichtssprache erklären (Kronig, 2000; Kronig, Haebelin, & Eckhart, 2000).

Dass die beiden Strukturkategorien „Nationalität“ und „soziale Herkunft“ nicht in loser Beziehung stehen, wird bei aufmerksamer Lektüre nicht entgangen sein: Mit der Übervertretung an „Fremdarbeiterknaben“ (Häfeli, 1979) waren explizit Knaben aus der „Unterschicht“ gemeint. Auch heute noch ist in Klassen des Schultyps mit Grundansprüchen mit einer überproportionalen Vertretung von Jugendlichen anderer nationalstaatlicher Zugehörigkeit (als der Schweizerischen) und einem höheren Anteil an Jugendlichen aus einer Familie mit niedrigem sozioökonomischen Status zu rechnen. Laut einer Analyse von Kronig (2007) erhält ein Schüler aus einer unterprivilegierten Zuwandererfamilie bei durchschnittlichen Schulleistungen mit einer Wahrscheinlichkeit von 51,7 Prozent eine Empfehlung in einen anspruchsvolleren Schultyp; bei einem mittleren bis hohen sozioökonomischen Status besteht für ihn hingegen eine Wahrscheinlichkeit von 59,3 Prozent. Ein unterprivilegierter Schweizer mit durchschnittlichen Schulleistungen kann von 61,6 Prozent ausgehen, ein Schweizer bei einem mittleren bis hohen sozioökonomischen Status gar von 83,3 Prozent.

Die mit einer Selektion angestrebte, jedoch bei weitem nicht eindeutig vorzunehmende Leistungsgruppierung hat neben den unmittelbaren Folgen auf die Klassenzusammensetzung auch weitreichende Folgen. Durch die Zuteilung in eine Sonderschule etwa wird bereits während der Primarschulzeit die zukünftige Ausbildungs- beziehungsweise Berufslaufbahn mit einer geringen Wahrscheinlichkeit auf Korrektur vorgespurt. So zeigten Auswertungen der amtlichen Bildungsstatistik des Kantons Zürich über zehn Jahre hinweg auf, dass nur rund fünf Prozent der Schülerinnen und Schüler aus Kleinklassen oder Sonderschulen in ein Angebot der Regelschule wechseln konnten (Gyseler, 2008). Während 70 Prozent der Schülerinnen und Schüler mit „besonderen pädagogischen Bedürfnissen“, welche integrativ in Regelklassen beschult wurden, eine weiterführende qualifizierende Ausbildung aufnehmen konnten, waren es bei denjenigen aus Kleinklassen und aus Sonderschulen nur bis zu etwa 50 Prozent. Auffallend war bei allen der drei untersuchten Gruppen von Schülerinnen und Schülern mit „besonderen pädagogischen Bedürfnissen“ der verzögerte Einstieg in eine Berufsausbildung – was vermuten lässt, dass die ursprünglichen Berufspräferenzen eher selten in Angriff genommen werden konnten.

Für Jugendliche mit Schulschwierigkeiten ist der Übergang von der Schule in den Beruf meistens und „seit jeher mit besonderen Hindernissen verbunden“ (Felkendorff & Lischer, 2005, S. 5). Steigende Ansprüche an Qualifikation, hohe Arbeitslosenquoten und zusätzliche Erschwernisse treffen heute und wohl auch in Zukunft Jugendliche mit niedrigeren Bildungsabschlüssen besonders. Bei einem Vergleich der Berufslaufbahnen von jungen Erwachsenen, welche ehemals als schulleistungsschwach galten und entweder integrativ oder separativ beschult wurden, ist „ein deutlicher Unterschied zugunsten der Integration“ (Riedo, 2000, S. 197) erkennbar. Diese aus narrativen Interviews gewonnenen Ergebnisse unterstützen die Aussage, dass mit dem Etikett „Sonderschülerin/Sonderschüler“ sehr geringe Chancen auf dem Ausbildungs- und Arbeitsmarkt verbunden sind (vgl. auch Ginnold, 2008; Gyseler, 2008; S. J. Wagner, 2005).

Die Selektion am Ende der Primarschulstufe bestimmt in ähnlicher Weise nicht nur über den Zugang zu einem bestimmten Schultyp der Sekundarstufe I. Gleichzeitig werden dadurch ebenfalls die Weichen für die spätere Bildungs- und Berufslaufbahn gestellt (u. a. T. Meyer, 2004; Neuenschwander, 2010, S. 15f.). Mit einem Schulabschluss des Schultyps mit Grundanforderungen sind – zunächst einmal grundsätzlich betrachtet – geringere Lehrstellenchancen als mit einem Schulabschluss des Typs mit erweiterten Anforderungen verbunden (Haeberlin, Imdorf, & Kronig, 2004; Imdorf, 2005). Überdies besteht vermehrt die Gefahr eines diskontinuierlichen Ausbildungsverlaufes (T. Meyer, 2004). Das Risiko, die postobligatorische Ausbildung schließlich ohne Abschluss zu beenden – oder mit anderen Worten: zu verlassen – ist für Absolventinnen und Absolventen des Schultyps mit Grundansprüchen deutlich höher als für solche aus Schultypen mit erweiterten Anforderungen (Gloor, Meier, & Nef, 2001).

Schulische Selektion verstärkt und bedingt Bildungsungleichheiten. Eine der Folgen sind unterschiedliche Klassenzusammensetzungen, welche innerhalb eines Schultyps der Sekundarstufe I gewisse Ähnlichkeiten aufweisen. Wie wirken sich die so gebildeten Referenzgruppen auf die Berufsabsichten und die weiteren Bildungsbestrebungen der Jugendlichen aus?

2.4.3 Referenzgruppeneffekte auf Bildungs- und Berufsaspirationen

Referenzgruppeneffekte spielen für den Berufswahlprozess eine nicht unerhebliche Rolle. Das durch die Referenzgruppe geformte akademische Selbstkonzept soll wichtige ausbildungsbezogene Entscheidungen einer Schülerin oder eines Schülers

zu beeinflussen vermögen – und zwar stärker noch als offizielle Leistungsmessungen wie Zensuren respektive Noten (Marsh & Craven, 2006; Marsh & Yeung, 1997b).

Köller (2004) etwa konnte einen Fischteich-Effekt auf das leistungsthematische Kurswahlverhalten von Gymnasialschülerinnen und -schülern ausmachen, der auch bei Kontrolle der Zeugnisnote nicht verschwand. Gemäß einer umfangreichen Studie von Ireson und Hallam (2009) hat das akademische Selbstkonzept gar einen größeren Einfluss auf die Absicht einer Schülerin oder eines Schülers in Zukunft weiterhin eine Schule zu besuchen beziehungsweise zu studieren als die gezeigten Schulleistungen. Von Bedeutsamkeit für das Schweizer Bildungssystem mit der von Kanton zu Kanton unterschiedlich stark gegliederten Sekundarstufe I könnte zudem das Hauptergebnis dieser Studie aus Großbritannien sein, wonach das Ausmaß der Leistungsgruppierung in einer Sekundarschule – nach Kontrolle der vorangegangenen Schulleistung und des früheren Selbstkonzepts – über die Höhe des akademischen Selbstkonzeptes, nicht aber über jene des generellen Selbstkonzepts mitbestimmt. In Schulen, in denen am häufigsten (d. h. in mehr als vier Unterrichtsfächern) nach Leistung gruppiert wurde, wiesen die Schülerinnen und Schüler im Durchschnitt ein tieferes akademisches Selbstkonzept auf, als in Schulen mit wenig Leistungsgruppierung. Die fachspezifischen Selbstkonzeptdomänen hatten überdies keinen Zusammenhang mit der Anzahl Schuljahre Erfahrung mit Leistungsgruppierung in Englisch, Mathematik und Naturwissenschaften – wohl aber mit der Position innerhalb der Leistungsgruppierungshierarchie. Die Autorinnen erklärten die Befunde anhand von Referenzgruppeneffekten, deren Referenzrahmen sowohl mit der Leistungsgruppe als auch mit deren Position innerhalb der Gruppierungshierarchie geboten sei (ebd.).

Dass eine „leistungsstarke“, hochselektive Schule neben Vorteilen auch Nachteile birgt, hatte bekanntlich bereits Davis (1966) festgestellt: Dazu veranlasste ihn sein Hauptbefund, wonach die Studenten in den Colleges mit einem hohem Anteil an leistungsstarken Studenten geringere Berufsaspirationen aufwiesen (vgl. Abschnitt 2.1). Im Jahre 1991 befand Marsh diese Konstatierung für untertrieben: Seine und weitere Ergebnisse würden belegen, dass solche *high-ability*-Schulen keine zusätzlichen akademischen Vorzüge gegenüber anderen Schulen bieten könnten. Als erstes wirke der erhöhte Leistungsdurchschnitt negativ auf das akademische Selbstkonzept, und dieses wiederum beeinflusse substantiell die Bildungsaspirationen. Der Leistungsdurchschnitt der Schule habe des Weiteren einen direkten negativen Einfluss auf die Berufsaspirationen, welches zu großen Teilen durch das akademische Selbstkonzept mediert sei.

Diese Befunde erscheinen vorerst wenig Aufmerksamkeit erregend. Das Besondere an der Untersuchung von Marsh (1991) war indes sein Ziel, eines der Desiderate aus dem Forschungsüberblick von Alwin und Otto (1977) zu bearbeiten:

die Identifizierung potenziell intervenierender Variablen des Schulkontextes auf Bildungs- und Berufsaspirationen (wie etwa bei J. W. Meyer, 1970). Entgegen ihren Annahmen konnten Alwin und Otto ihrerzeit keinen substantiellen Effekt der beiden Kontextmerkmale Leistungsdurchschnitt sowie mittleren sozioökonomischer Status einer Schule auf die Bildungs- und Berufsaspirationen ausmachen, wohl aber einen kleinen negativen *frog pond effect* (vgl. Davis, 1966).

In Bezug auf den mittleren sozioökonomischen Status äußerte sich Marsh damals (1991) ähnlich wie vier Jahre zuvor: Der BFLPE hänge grundsätzlich nicht von dessen Inklusion oder Exklusion ab. Interessant wäre allerdings zu ergründen, warum die beiden Schulkontextvariablen differente Effekte hätten. Einerseits könnten die negativen Effekte aufgrund sozialer Vergleichsprozesse zustande kommen, die positiven andererseits aufgrund von Gruppenidentifikation. Marsh verwies zudem auf die Schwierigkeit der Entflechtung der beiden Effekte aufgrund der substantiellen Korrelation zwischen dem Leistungsdurchschnitt und sozioökonomische Status einer Schule.

Einhergehend mit der Entwicklung von neuen statistischen Analysemöglichkeiten, führte Marsh gemeinsam mit O'Mara (2010) eine Reanalyse der breit angelegten, acht Jahre und fünf Messzeitpunkte umfassenden US-amerikanischen Studie „*Youth in Transition*“ anhand von Strukturgleichungsmodellen unter Berücksichtigung der Gruppenabhängigkeit der Individualvariablen durch. Einer der Hauptbefunde ist, dass der negative direkte Effekt des schulinternen Leistungsdurchschnitts auf das individuelle akademische Selbstkonzept – entspricht dem BFLPE – nach der obligatorischen Schulzeit nicht nur bestehen blieb, sondern im Verlaufe der akademischen und beruflichen Laufbahn kumulierte. Ebenfalls persistierend negativ wurden die Bildungsbestrebungen sowie die beruflichen Aspirationen durch den schulinternen Leistungsdurchschnitt beeinflusst. Diese Ergebnisse sprechen für eine Erweiterung des BFLP-Effekts auf Bildungs- und Berufsaspirationen. Des Weiteren zeigte sich, dass der negative Effekt des Leistungsdurchschnitts einer Schule auf die Zielvariablen durch das akademische Selbstkonzept mediiert wurde.

Besonders aufschlussreich ist außerdem – gerade auch vor dem Hintergrund der vorgängig angetönten Debatte (Alwin & Otto, 1977; Marsh, 1987, 1991) – die zusätzliche Berücksichtigung des mittleren sozioökonomischen Status einer Schule sowie des individuellen sozioökonomischen Status. Bei den Analysen zeigte sich bei beiden Merkmalen ein positiver Effekt sowohl auf das akademische Selbstkonzept als auch auf die Bildungs- und Berufsaspirationen. Allerdings vermochte der positive Effekt des mittleren sozioökonomischen Status einer Schule den negativen Effekt des Leistungsdurchschnitts der Schule nicht aufzuheben. Somit unterstützen die Resultate einerseits die bisherigen theoretischen Annahmen des BFLP-Effekts, wonach das Leistungsniveau einer Schule die Schlüsselrolle einnimmt. Andererseits

veranlassten die Ergebnisse Marsh und O'Mara (2010) zu Präzisierungen des BFLPE-Modells: Der BFLPE entspricht dem Nettoergebnis (*net effect*) eines großen negativen Kontrasteffektes und eines um einiges kleineren positiven Assimilationseffektes. Des Weiteren ist die Kontrolle des individuellen sozioökonomischen Status notwendig.

Trotz der hohen Korrelation der beiden Schulkontextvariablen stellen der BFLPE und der Assimilationseffekt – Marsh und O'Mara sprechen an dieser Stelle auch von *school climate effect* – zwei unterschiedlichen Phänomene dar. Diese Juxtaposition zwischen mittlerem sozioökonomischem Status und Leistungsniveau einer Schule, zwischen Assimilation und Kontrastierung, betrachten sie als ein wichtiges zukünftiges Forschungsthema.

2.5 Kontroverse zum referenzgruppentheoretischen Modell nach Marsh

Trotz seiner sehr guten empirischen Absicherung (z. B. Seaton, Marsh, & Craven, 2009) und seiner Popularität ist der Referenzgruppeneffekt nach Marsh beziehungsweise der *big-fish-little-pond effect* (BFLPE) nicht unumstritten. Kritisiert werden in erster Linie die theoretische Konzeptualisierung sowie inhaltliche Schlussfolgerungen hinsichtlich praktischer Implikationen.

Ein erster Diskussionspunkt betrifft die grundlegende Frage, ob es sich beim BFLPE um eine Theorie, ein Modell oder schlicht um einen Effekt handelt. Bei einer eng gefassten Definition stellt der BFLPE lediglich das besagte häufig vorzufindende Ergebnismuster dar. Für eine Einordnung als Theorie wären neben einer Beschreibung und Erklärung des Phänomens, dessen Auftretensbedingungen sowie Vorhersagen zu den damit assoziierten Prozessen aufzuzeigen. Die logisch konsistente Verknüpfung von gut bewährten und anerkannten Hypothesen bestimmt letztendlich den Kern von sozialwissenschaftlichen Theorien (Bortz & Döring, 2006). Während der BFLPE durchaus Merkmale einer Theorie aufweist – die Beschreibung eines Phänomens und die Vernetzung von einzelnen Aussagen zu einem zusammenhängenden Ganzen, fehlt indessen gemäß Trautwein und Lüdtke (2010) „eine allgemein akzeptierte, wissenschaftlich exakte Formalisierung der Annahmen“ (S. 13). Nichtsdestotrotz werden in der Forschungspraxis gewisse Annahmen über Bedingungen oder Mechanismen meist mehr oder weniger explizit mit dem BFLPE in Verbindung gebracht.

Im Grunde bediente sich Marsh (1987) eines bildlichen Vergleichs, einer Metapher, um einen bisher unbekannten Gegenstandsbereich auf einen bekannten

Sinngehalt zu übertragen. Durch das Bild des großen Fisches im kleinen Teich wurde das verständlich, was sich mit keinem bereits vertrauten Begriff bestimmen ließ. Die Bildung von Metaphern erhält wissenschaftstheoretisch betrachtet eine heuristische Bedeutung. Sie gelten als Modelle – ebenso wie die Interpretation von mathematischen Strukturen beispielsweise in der Statistik, die der empirischen Forschung einen ontologischen Rahmen bieten (Herzog, 2012). Nach Hübner (1982) kann ein Modell als „apriorische Konstruktion“ (S. 271) verstanden werden. Eine solche Strukturierung beeinflusst gleichzeitig die Theoriebildung (Beck & Krapp, 2006). Anbetracht dieser in groben Zügen umrissenen Überlegungen lässt sich der BFLPE als Modell mit einem referenztheoretischen Hintergrund einordnen.

Die mangelnde theoretische Konzeption des BFLP-Effekts bietet den Kritikerinnen und Kritikern eine vortreffliche Diskussionsplattform. Dai (2004; 2008) etwa warf der Forschung rund um den BFLPE einen minimalen Kontakt zur Fachliteratur sozialer Vergleichstheorien vor. Als ein Hauptproblem des BFLPE-Modells erachtete er die allzu starke Vereinfachung von weit komplexeren Prozessen. So wird im Fachgebiet sozialer Vergleichstheorien beispielsweise davon ausgegangen, dass – entgegen der These des Fischteich-Effektes – soziale Vergleiche in unterschiedlichen Richtungen vorgenommen werden: Je nach Kontext, Personmerkmale oder Ziel seien Vergleiche mit leistungsmäßig ähnlichen Mitschülerinnen oder Mitschülern, aufwärts mit leistungstärkeren oder abwärts mit leistungsschwächeren möglich (vgl. auch Banaji & Prentice, 1994; Collins, 1996; Wood, 1996). Sowohl Aufwärts- als auch Abwärtsvergleiche können – in Abhängigkeit von Situation oder Kontext – positive oder negative Konsequenzen nach sich ziehen (Lockwood & Pinkus, 2008). So vermag etwa der Vergleich mit einer Person, der es schlechter zu gehen scheint, Auftrieb nach einer missglückten Erfahrung zu geben. Im Gegenzug kann ein Abwärtsvergleich entmutigen, wenn dabei die Aussicht auf eine ähnlich düstere Zukunft gewahrt wird. Aufwärtsvergleiche dienen laut Taylor und Lobel (1989) dazu, den Selbstwert zu verbessern; demgegenüber unterstützen Abwärtsvergleiche eine Person bei der positiven Bewältigung einer Situation und erhöhen gleichzeitig die Motivation. Des Weiteren wird je nach Richtung der sozialen Orientierung eine andere Wirkung auf die Leistung erwartet (Wehrens et al., 2010).

Verschiedentlich wurde zudem darauf hingewiesen, dass soziale Vergleiche auf verschiedenen Ebenen vorgenommen werden können: Einerseits innerhalb einer von außen oder institutionell vorgegeben Referenzgruppe; andererseits in einem selbst gewählten sozialen Bezugsrahmen – beispielsweise mit leistungsstarken Gleichaltrigen aus anderen Schulklassen was leistungsförderlich wirken soll (Blanton, Buunk, Gibbons, & Kuyper, 1999). Ob eine Person sich mit anderen vergleicht (Kontrasteffekt) oder sich an jemanden angleicht (Assimilationseffekt), sei abhängig von der Art der Selbst-Aktivierung (Stapel & Koomen, 2001). Gemäß

McFarland und Buehler (1995) sind außerdem Personen, die ihre Bezugsgruppe positiv einschätzen, dem BFLPE nicht ausgesetzt. So schließen sie aus den Ergebnissen ihrer experimentellen Studien, dass zwar „everyone feels good about being a big fish in a little pond“, aber „not everyone feels bad about being a little fish in a big pond“ (ebd., S. 1068).

Ein wiederholt gesetztes Fragezeichen betrifft den in Forschungsdesigns verwendeten Referenzrahmen für den BFLPE. In den bisherig aufgeführten Studien zum BFLPE von Marsh und Mitarbeitenden wurde generell als Vergleichsstandard der Leistungsdurchschnitt innerhalb einer Schule definiert. Damit ist implizit die Annahme verknüpft, alle Schülerinnen und Schüler einer Schule würden ihre sozialen Vergleiche anhand eines gemeinsamen, schulbasierten Vergleichsmaßstabes vornehmen. Dai und Rinn (2008) geben in ihrer ausführlichen Auseinandersetzung – die nota bene von Marsh begutachtet wurde – als einer der Hautkritikpunkte an, Untersuchungen zum Fischteich-Effekt würden keine direkten Belege dafür bieten, dass auch tatsächlich soziale Vergleiche innerhalb einer Schule das Phänomen steuern. Außerdem sei der Kontext in den meisten Studien zu unspezifisch. Neben dem sozioökonomischen Status, der sozialen Herkunft oder der Klassengröße ordneten sie die Intensität an Leistungsgruppierung als weitere wichtige Kontextvariable ein. Der letztgenannte Kritikpunkt fand durch die Ergebnisse der Studie von Ireson und Hallam (2009) Unterstützung (vgl. Abschnitt 2.4.3).

Eine Antwort lieferte kurz darauf eine Publikation einer Autorengruppe rund um Huguet (2009) – wozu auch Marsh zählte. In der in Frankreich durchgeführten Studie wurden über 2000 Schülerinnen und Schüler aus 99 Schulklassen während ihrem ersten Sekundarschuljahr unter anderem zu ihrem akademischen Selbstkonzept, zu der von ihnen wahrgenommenen relativen schulleistungsbezogenen Position innerhalb der jeweiligen Schulklasse sowie zu den von ihnen für Leistungsvergleiche bevorzugten Klassenkolleginnen und Kollegen befragt. Die Kontrolle der relativen Position innerhalb der Schulklasse hatte eine Elimination des Fischteich-Effektes zur Folge, was als eindeutiger Anhaltspunkt für die postulierten sozialen Vergleiche mit der Schulklasse als Ganze gewertet wurde. Mit Anstieg des Leistungslevels in einer Klasse wurden diese sozialen Vergleiche von Jugendlichen überdies missgünstiger und unfairer erlebt.

Ein weiterer erwähnenswerter Befund der Untersuchung von Huguet und Mitarbeitenden ist der außergewöhnlich hohe BFLPE-Koeffizient (von -0.46), welcher damit erklärt wurde, dass die Schulklassen den unmittelbareren Referenzrahmen darstellen würden (üblicherweise ist mit einem Koeffizienten von -0.20 zu rechnen, vgl. Marsh & Hau, 2003). Darüber hinaus zeigte sich eine Interaktion zwischen dem BFLPE und individuellem Leistungsniveau: Zwar konnte wie erwartet der BFLPE für alle Leistungslevels ausgemacht werden (wie etwa bei Köller et al.,

2000), für Schülerinnen und Schüler mit schwachen Leistungen war dieser Effekt jedoch stärker als für die Schülerinnen und Schüler mit guten Leistungen (Huguet et al., 2009). Bis anhin waren entsprechende Interaktionseffekte meist gering und nicht signifikant, und falls doch, nicht konsistent (Marsh, Seaton, et al., 2008). Als Erklärung diente ebenfalls der enger als üblicherweise gewählte Referenzrahmen.

Verschiedentlich angezweifelt wurde außerdem die Übertragbarkeit des BFLP-Effekts auf die hochselektiven Bildungssysteme der Schweiz oder von Deutschland. Trautwein und Lüdtke (2005) äußerten sich etwa dahingehend, es sei unklar, inwieweit gewisse Schulstrukturen das Verhalten der Lehrperson und gewisse Charakteristiken der Schülerschaft den BFLPE verstärken oder eindämmen würden. In Frage stellten sie den negativen Referenzgruppeneffekt auf die Bildungs- und Berufaspirationen, da in einem stark differenzierenden Schulsystem die Qualifikation eines anspruchstieferen Schultyps zu etlichen weiterführenden Ausbildungen den Zugang verunmöglichen würden. Worauf Marsh (2005) antwortete, dass die bestehenden Studien die Robustheit des BFLP-Effekts gut belegen würden (vgl. hierzu die 26 Länder übergreifende Studie von Marsh & Hau, 2003), es aber möglicherweise gewisse Situationen und Settings geben würde, für deren Erklärung der BFLPE nicht herangezogen werden könne. Marsh (2005b) befürwortete im Übrigen die Mehrheit der von Trautwein und Lüdtke angebrachten Vorschläge – was in mehreren darauffolgenden gemeinsamen Forschungsarbeiten Ausdruck findet (u. a. Marsh, Trautwein, Lüdtke, & Köller, 2008; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006; Trautwein, Lüdtke, Marsh, & Nagy, 2009).

Weniger einverstanden erklärte sich Marsh mit der in demselben Themenheft erschienenen Stellungnahme von Kornmann (2005): Sie biete weder rationale noch professionelle Erklärungen zur eingebrachten Behauptung, warum die Beschulung von Kindern und Jugendlichen mit „milden intellektuellen Lernschwierigkeiten“ in Regelklassen zu besseren Leistungen führen solle – statt zu demoralisierenden Erlebnissen, welche doch eigentlich effektives Lernen untergraben würden (Marsh, 2005b). Kornmanns Verweis auf die Ergebnisse der Studie von Haeberlin und Mitarbeitern (1990) – bezüglich der größeren Leistungsentwicklung in integrativen Klassen im Vergleich zu separativen Klassen – schien für Marsh nicht von Bedeutung zu sein. Zumindest fanden die beiden Autoren Übereinstimmung zum einen in dem Punkt, dass der negative Referenzgruppeneffekt auf das Fähigkeits-selbstkonzept respektive auf das akademische Selbstkonzept ein seriöses Problem darstelle, welches in der sonderpädagogischen Forschung und Praxis eingehender behandelt werden sollte. Zum anderen teilten sie die Überzeugung, dass die Förderung eines „realistischen“ und zugleich optimistischen Selbstkonzepts möglich sei. Dazu seien die eigens für die Bedürfnisse dieser leistungsschwachen Schülerinnen

und Schüler konzipierten Sonderklassen am besten geeignet – wie Ergebnisse aus ihren Untersuchungen belegen würden (Marsh, 2005b).

Aus diesem als letzten aufgeführten Diskussionspunkt wird erneut ersichtlich, dass die Kontroverse rund um den BFLPE unweigerlich Fragen rund um Wirkungsweisen von Selektion und Leistungsgruppierung mit einschließt. Gemäß dem Referenzgruppeneffekt nach Marsh wird Kindern und Jugendlichen mit schwachen Schulleistungen in separierenden Schulklassen oder in Klassen mit Grundansprüchen aufgrund des tieferen Leistungsdurchschnittes der sozialen Bezugsgruppe eine bessere Entwicklung des akademischen Selbstkonzepts als in Regelklassen oder in Klassen mit erweiterten Ansprüchen vorausgesagt. Hauptargument für die leistungsorientierten Homogenisierungsbestrebungen im Rahmen von selektiven Bildungssystemen ist allerdings eine optimale schulische Leistungsentwicklung – nicht eine möglichst hohe Selbsteinschätzung der schulischen Leistungsfähigkeit (vgl. Kritik von Plucker et al., 2004). Auf Seiten der Proponenten des komparativen Effekts der schulischen Bezugsgruppe wird daher die Bedeutung des akademischen Selbstkonzepts gleichzeitig mit seiner Rolle als Kriteriums- als auch als Prädiktorvariable hervorgehoben. Mögliche Vorteile der gemeinsamen Beschulung mit leistungsstärkeren Klassenkameradinnen und Klassenkameraden oder eines Unterrichts mit höheren Ansprüchen werden im theoretischen Modell des BFLP-Effekts nicht – oder mit dem Assimilationseffekt nur ansatzweise – berücksichtigt.

Trautwein und Lüdtke (2005) verweisen denn auch nicht ohne Grund darauf hin, neben der hitzigen und ungelösten Debatte über die Struktur und Größe des Teiches die Qualität des „Futters“ nicht zu vergessen. Schließlich sei dieses in erster Linie verantwortlich für die Entwicklung des Fisches.

Schulklasseneffekte

Schülerinnen und Schüler zwischen komparativen und
normativen Einflüssen

Zurbriggen, C.

2016, XVII, 274 S. 22 Abb. in Farbe., Softcover

ISBN: 978-3-658-13139-5