

Super Title

Master Thesis

David-Matthias Sichau

Supervisors:  
Pitt Hild,

GroupName

24. September 2014, Zürich

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theorie und Problemlage</b>	<b>2</b>
2.1	Transfer	2
2.1.1	Historischer Überblick	2
2.1.2	Kritiken	4
2.1.3	Lobato und Sievert 2002	4
2.1.4	Elemente von Transfer	5
2.1.5	Konsequenzen für den Unterricht	6
2.1.6	Zusammenfassung zu Transfer	7
2.2	Kompetenz	8
2.2.1	Bildungsreformen	8
2.2.2	Definition von Kompetenz	9
2.2.3	Kompetenz und Transfer	10
2.3	Kompetenz des skalenbasiertes Messens	11
2.4	Forschungsfrage	11
<b>3</b>	<b>Untersuchungsanlage</b>	<b>13</b>
3.1	Anforderungen	13
3.2	Umsetzung	13
3.3	Durchführung	14
3.3.1	Vorbereitung	14
3.3.2	Durchführung	15
3.3.3	Nachbereitung	15
<b>4</b>	<b>Ergebnisse</b>	<b>17</b>
4.1	Kodierung	17
4.1.1	Items	17
4.1.2	Qualitätsstandards	18
4.1.3	Niveau	19
4.2	Fragebogen	20
4.3	Unterschiede zwischen den Klassen	21
<b>5</b>	<b>Diskussion</b>	<b>22</b>
<b>6</b>	<b>Ausblick</b>	<b>23</b>
	<b>Literaturverzeichnis</b>	<b>24</b>
	<b>Anhang</b>	<b>28</b>

---

# Abstrakt

Code erhältlich auf:

**GitHub**

<http://sichau.eu>

Toller Abstrakt

---

# 1 Einleitung

Im Rahmen der ExKoNawi hands-on Testaufgaben soll untersucht werden inwiefern die Qualitätsstufen von verschiedenen Test zur Erhebung der Kompetenz des Messens korrelieren.

Dabei soll insbesondere untersucht werden, ob der fachliche Kontext oder der inhaltliche Kontext ausschlaggebend ist. In anderen Wort hängt die Leistung der SuS davon ab, ob die Kompetenz des Messens im gleichen Fach gemessen wird, oder hängt dies davon ab was für eine Messung durchgeführt wird.

Diese Frage ist besonders von Interesse, wenn es darum geht Test auszuwählen. Ist es möglich die Kompetenz des Messens mit einem einzelnen Test zu erheben oder sind mehrerer Tests notwendig um das erreichte Kompetenz Niveau ausreichend zu beschreiben.

---

## 2 Theorie und Problemlage

Für die Beantwortung und Entwicklung der Forschungsfrage ist es wichtig den Begriff des Kontextes zu definieren. Zu Beginn wird der Begriff des Kontextes aus dem Sichtwinkel des Transfers untersucht. In einem zweiten Teil wird der Kontext basierend auf dem Kompetenzbegriff analysiert. Zuletzt werden die Erkenntnisse gesammelt und die genaue Forschungsfrage dieser Masterarbeit definiert.

### 2.1 Transfer

Es wird erwartet, dass in der Schule vermitteltes Wissen universell aufgerufen werden kann und dass das Gelernte auf andere Kontexte angewendet werden kann. Dieses universell verfügbare Wissen ist eng mit dem Begriff des Transfers verknüpft. Greeno, Collins und Resnick (1996) definierten Transfer als „the process of applying knowlege in new situations“. Aber auch innerhalb der schulischen Bildung gibt es einen Transfer zwischen den verschiedenen Fächern. So wird von Schülerinnen und Schüler erwartet, dass Fähigkeiten, Lösungsstrategien, Konzepte und anderes Wissen auf andere Fächer übertragen werden soll und dort abgerufen werden kann.

#### 2.1.1 Historischer Überblick

Um den Begriff des Kontextes im Zusammenhang mit Transfer zuverorten, soll zuerst ein historischer Überblick über den Begriff des Transfers gegeben werden.

##### Woodworth 1901

Eines der ersten Experimente zu Transfer wurde von Woodworth und Thorndike (1901) gemacht. Dabei mussten Probanden die Grösse von Rechtecken schätzen. Nachdem die Personen sich durch Wiederholungen verbessert hatten, wurde ihnen zwei neue Test Sets gegeben. In einem gab es neue Rechtecke, welche im ursprünglichen Set nicht enthalten waren. Die zweite Gruppe bekam Sets bei denen andere Formen enthalten waren (z. B. Kreise und Dreiecke). Die zweite Testgruppe machte ähnlich viel Fehler, wie vor dem Training mit den Rechtecken. Daraus schloss Woodworth und Thorndike, dass kein Transfer stattgefunden haben kann.

Ein ähnliches Resultat auf universitärem Niveau konnte von Renkl et al. (1994) gezeigt werden. Er konnte zeigen, dass Nichtökonominnen eine simulierte Firma besser führten, als Studierende der Betriebswissenschaften kurz vor ihrem Abschluss. Diese Resultate führen zu dem Schluss, dass Transfer nur sehr schwierig erreicht werden kann, und wenn oft nur unter sehr ähnlichen Bedingungen. Die Woodworth und Thorndike Theorie zu Transfer basiert auf der Idee von identischen Elementen (Pea 2013). In diesem Theorie-Verständnis entsteht Transfer, wenn Wissen auf zwei verschiedenen Aufgaben, welche jedoch identische

---

Merkmale/Elemente besitzen, angewendet wird. Dieses Transfers Verständnis basiert und stützt das Reiz-Reaktions-Modell des Lernens (Detterman 1993; Mietzel 2007).

### **Ferguson 1956**

Eine alternative Theorie zum Transfer wurde von Ferguson (1956) entwickelt. Fergusons Theorie basiert darauf, dass die Intelligenz einer Person sich auf deren Transferleistung auswirkt. So findet nach Ferguson (1956) bei dem Lernen permanent ein Transfer statt, da jede Lernaufgabe von der anderen unterschiedlich ist und daher Transfer stattfinden muss. Im Unterschied zu Woodworth und Thorndike (1901) betrachtet Ferguson Transfer als einen kontinuierlichen Prozess, welcher durch Lernen verbessert werden kann. Wichtig ist jedoch zu beachten, dass Ferguson Theorie nur Nah-Transfer beschreibt. Unter Nah-Transfer wird Transfer zwischen sehr ähnlichen Situationen definiert.

### **Judd 1908**

Eine der grundlegenden Studien zu Fern-Transfer, bei welchem erworbenes Wissen auf Kontexte angewendet werde soll welche sich deutlich vom Kontext, unter welchem das Wissen erworben wurde, unterschieden, wurde von Judd (1908) gemacht. Im Vergleich zu Woodworth und Thorndike geht Judd davon aus, dass der Unterschied zwischen den beiden Situationen nicht nur abhängig von der Ähnlichkeit und den Unterschieden zwischen den beiden Situationen ist, sondern auch davon abhängt, wie die erste Situation gelernt wurde. Um dies zu belegen führte Judd eine sehr bekannte Studie durch. Bei dieser wurden Kinder genommen, welche mit einem Dart auf eine Zielscheibe unter Wasser werfen sollten. Beide Gruppen bekamen zu beginn die Möglichkeit dies zu trainieren. Später wurde das Werfen wiederholt, wobei die Position der Zielscheibe jedoch unterschiedlich war, und untersucht welche Gruppe besser war. Eine der beiden trainierten Gruppen wurde währen sie die Situation A trainierten erklärt, warum die Scheibe so schwierig zu treffen war. Indem ihnen zusätzlich zum Training noch das Prinzip der Lichtbrechung erklärt wurde. Die Gruppe welche die Erklärung bekommen hatte schnitt unter der neuen Situation deutlich besser ab, als die andere Gruppe. Judd (1908) erklärte dieses damit, dass die einen wussten, welches Prinzip sie auch bei der zweiten Situation anwenden können. Die Schülerinnen und Schüler welchen das Prinzip nicht erklärt wurden, haben gelernt ihren Wurf auf die erste Situation anzuwenden, konnten dieses Wissen jedoch nicht generalisieren, da dies spezifisch für die Situation erworben wurde.

Im Vergleich zu Woodworth und Thorndike beinhaltet die Theorie von Judd einen kognitivistisches Verständnis des Lernens. Da die Lernenden ein immer besseres Verständnis der Welt um sich selbst konstruieren und so neue Situation basierend auf ihrer internen Repräsentation der Welt lösen können. Detterman (1993) kritisiert an dieser Studie die Verwendung von Transfer. So erklärt Judd einem Teil der Personen das zugrunde liegende Prinzip. Dies ist laut Detterman jedoch äquivalent, wie wenn man den Personen sagen würde, dass sie dieses Prinzip verwenden sollen. Was dann identisch wäre, wie wenn man einer Anleitung folgen würde.

---

## Gick und Holyoak 1980

Eine weitere bedeutende Studie zu Transfer wurde von Gick und Holyoak (1980) durchgeführt. Dabei wurde untersucht, unter welchen Bedingungen Lernende Analogien verwenden, um strukturell ähnliche Probleme zu lösen. Ein Beispiel Problem welches sie den Lernenden gaben war, wie kann ein Tumor mit Strahlung zerstört werden, ohne dass gesundes Gewebe geschädigt wird. Dieses Problem wurde erstmals von Duncker und Lees (1945) verwendet. Dieses Problem kann gelöst werden, indem man mehrere Strahlen verwendet, welche sie nur im Tumor überlagern. Bevor sie dieses Problem lösen bekommen die Lernenden eine Geschichte erzählt, bei welcher das gleiche Prinzip verwendet wird. In dieser Geschichte ging es darum ein Fort, welches von Minen umgeben ist zu erobern. Durch aufteilen der Angreifer in mehrere angreifende Gruppen, die unterschiedliche Wege gehen, wurde die Belastung auf die Minen reduziert und das Fort konnte erobert werden. Das Resultat dieser Studie zeigte, dass spontaner Transfer nur sehr selten stattfindet. Das Hören der Geschichte führt nicht zu einer höheren Wahrscheinlichkeit das zweite ähnliche Problem zu lösen, solange die Lernenden nicht auf die Ähnlichkeit aufmerksam gemacht werden.

### 2.1.2 Kritiken

Lave (1988) kritisiert die verschieden hier vorgestellten Untersuchungen. Da bei allen angenommen wird, das Wissen automatisch generalisierbares Wissen erzeugt, welches auf verschiedene Situationen angewendet werden kann. Sie schlägt eine Alternative vor welche sie als “practice view” bezeichnet. Bei dieser wird Wissen von Personen erworben, welche an speziellen Übungen teilnehmen und daraus nur Wissen entwickelt wird, welches auf diese spezifische Situation(Kontext) zutrifft.

Folgende Kritiken erhebt sie. So stellt sie die Frage was lernen die Teilnehmer der verschiedenen Studien überhaupt. So greift sie insbesondere die Annahme an, dass die Teilnehmer der Studien Kontext unabhängig lernen. Sie lernen immer Kontext spezifisch. Ein anderer Punkt welche Sie angreift ist, wer definiert die Ähnlichkeit der Probleme. Ist die Ähnlichkeit der Probleme für die Teilnehmer der Studie auch greifbar. Auch Detterman (1993) kritisiert die Studien. So sollten seiner Meinung alle Studien zu Transfer als Doppel-Blind Studien durchgeführt werden, da der Studienleiter unbewusst die Leistung der Probanden ändern kann. Detterman fordert daher:

No tranfer experiment should be carried out without using a double blind procedure, particularly experiments assessing general transfer Detterman (1993, S. 10).

### 2.1.3 Lobato und Sievert 2002

Nachdem einige historische Studien zu Transfer exemplarisch aufgezeigt wurden, soll eine aktuelle Studie zu Transfer welche auf die Kritiken eingeht gezeigt werden.

Lobato und Siebert (2002) möchte einen Kritik-Punkt von Lave (1988) lösen. So kritisierte Lave, dass der Untersucher festlegt was Transfer von Wissen ist. Daher legten Lobato

und Siebert (2002) als Messung für Transfer fest, welche Ähnlichkeit der Proband selbst zwischen verschiedenen Situationen zieht. So untersuchten Sie einen Schüler, welcher eine Rollstuhlrampe erhöhen sollte ohne die Steigung zu verändern. Der Schüler löste dieses Problem in dem er die Verhältnisse von Höhe zu Länge konstant hielt. Er verwendete dafür jedoch nicht die im Mathematik Unterricht gelernten Formeln. In den bisherigen Untersuchungen wäre daher angenommen worden, dass der Schüler keinen Transfer geleistet hat. Aufgrund der Interviews stellte sie jedoch fest, dass der Schüler sehr wohl Transfer geleistet hat, indem er das Konzept von konstanter Geschwindigkeit als Verhältnis von zurückgelegter Strecke zur Zeit auf dieses Problem angewendet hatte.

Lobato und Siebert (2002) konnten damit zeigen, dass wenn man die Ähnlichkeit zwischen zwei verschiedenen Situationen(Kontexten) nicht mit strukturellen Ähnlichkeiten oder Unterschieden beschrieben werden sollen. Sondern damit, wie der Lernende die Ähnlichkeiten zwischen den Situationen(Kontexten) wahr nimmt.

### 2.1.4 Elemente von Transfer

Nachdem ein Überblick über die historische Entwicklung von Transfer gegeben wurde, soll nun auf die grundlegenden Elemente, welche bei Transfer anzutreffen sind eingegangen werden.

Marini, McKeough und Lupart (1995) definieren drei Elemente, welche zu einem Transfer führen. Das erste Element besteht aus Merkmale des Lernenden. Dieser hat sobald er eine Situation antrifft bereits ein bestimmtes prozedurales und deklaratives Wissen, welches er sich erarbeitet hat und abrufen kann. In einem bestimmten Kontext kann er einen Teil davon abrufen und anwenden (s. S. 189ff). Dies führt dazu das lösungsrelevantes Wissen von den Vorhanden und dem verarbeitbarem Wissen abhängt. Zusätzlich kann in einem bestimmten Kontext jedoch nicht alles Wissen abgerufen werden, da man mit tragem Wissen rechnen muss und auch der aktuellen Motivation des Lernenden.

Als zweites Element von Transfer gibt Marini, McKeough und Lupart die Merkmale einer Aufgabenstellung an. So hängt Transfer von der Ähnlichkeit der Aufgabe ab. Dabei gibt es jedoch einen Unterschied zwischen Novizen und Experten. Novizen vergleichen Aufgaben hauptsächlich aufgrund oberflächlicher Merkmale, wohingegen Experten sich auf die zugrunde liegenden Prinzipien fokussieren (s. S. 279). Aufgrund dessen, haben Novizen oft Problem den Zusammenhang zwischen Aufgaben zu sehen und können daher keinen Transfer durchführen.

Das dritte Element ist der Kontext in den ein Problem eingebettet ist. Ein Beispiel dafür ist die Untersuchung von Godden und Baddeley (1975). Dort lernten Taucher Wörter Unterwasser auswendig. Bei einer späteren Überprüfung konnten sie sich an mehr Wörter erinnern, wenn es Unterwasser wiederholt wurde im Vergleich zu einer Wiederholung auf dem Festland. Dieser Ortswechsel ist auch bei ausserschulischem Kontext gegeben. Aber auch innerhalb der Schule kann es zu unterschieden kommen. Ein Beispiel dafür liefert Schoenfeld (1988) so hatten Lernende keine Schwierigkeiten mit einer Divisionsaufgabe. Wenn die Aufgabe jedoch in einen Kontext gestellt wurde, wie z.B. in eine Textaufgabe eingebettet wurde, scheiterten die meisten der Lernenden.

Erst durch die Berücksichtigung aller drei Elemente lässt sich Transfer ganzheitlich Be-



trachten. Nicht wie Woodworth und Thorndike (1901), welcher nur den Aspekt der Aufgabenmerkmale genauer untersucht hat. Erst neuere Arbeiten berücksichtigen alle Elemente und insbesondere den Kontext (Lobato und Siebert 2002; Detterman 1993; Greeno, Collins und Resnick 1996)

### 2.1.5 Konsequenzen für den Unterricht

Wie Claxton (1990) zeigte, darf jedoch nicht davon ausgegangen werden, dass in der Schule erworbenes Wissen ohne weiteres auf andere Alltags Probleme angewendet werden können. So sprach Whitehead (1929) von „trägem Wissen“ (inert knowledge), wenn Wissen vorhanden ist um ein Problem zu lösen, dieses jedoch nicht automatisch abgerufen werden kann. Dieses Wissen ist erst greifbar, wenn die Person angeregt wird dieses Wissen zu verwenden. Nach Whitehead (1929) entsteht träges Wissen oft unter schulischen oder universitären Bedingungen.

Detterman geht sogar noch weiter und schliesst aus den Studien zu Transfer:

that, if you want people to learn something, teach it to them. Don't teach them something else and expect them to figure out what you really want them to do (Detterman 1993, S. 21).

Andere Autoren haben jedoch keine so pessimistische Sicht auf die Fähigkeit zu Transfer und geben Empfehlungen, wie schulischer Unterricht aussehen muss, welcher verhindert, dass träges Wissen entsteht und möglichst viel Transfer von Wissen stattfinden kann.

### Überlernen von Fähigkeiten

Eine Möglichkeit, gute Transferleistung zu erreichen, ist das intensive einüben von Grundfertigkeiten, wie zum Beispiel in der Grundschule. LaBerge und Samuels (1974) untersuchten dies bei der Fertigkeit des Lesens. Dabei wird das Üben nicht abgebrochen, wenn die Schülerinnen und Schüler die Fertigkeit subjektiv bereits können, sondern noch einige Zeit fortgesetzt. LaBerge und Samuels haben dabei Schülerinnen und Schüler einen Text solange laut vorlesen lassen, bis sie keinen Fehler mehr machten und einen hohen Flüssigkeitsgrad aufwiesen. Dieses *überlernen* einer Fertigkeit fördert Transfer. So führt nach Perkins und Salomon (1989) hochgradig eingeübte Fertigkeiten zu spontanem automatischem Transfer, ohne dass es längeren Nachdenkens bedarf. Der Grund dafür liegt darin, dass Routinen gebildet wurden, welche in einer neuen Situation helfen, die Aufmerksamkeit verstärkt auf neue Aspekte zu richten (LaBerge und Samuels 1974; Mietzel 2007).

Diese Erkenntnis deckt sich mit den Forderungen von Whitehead (1929), welcher bereits 1929 davor warnte, dass in der Schule träges Wissen entsteht. Daher soll in der Schule darauf geachtet werden, nicht zu viel in zu kurzer Zeit zu erreichen. So fordert er auch wenige Themen gebiete gründlich zu erarbeiten. Diese Forderung wurde auch von neueren Studien bestätigt (Porter 1989; Brophy 1992; Millar und Osborne 1999).

---

## Entkontextualisieren

Wie bereits vorher angesprochen, hängt Wissen sehr stark vom Kontext ab unter welchem es gelernt wurde (Godden und Baddeley 1975; Schoenfeld 1988). Anderson, Reder und Simon (1996) fordern daher, dass Wissen, so erworben werden soll, dass Lernende lernen, irrelevante Aspekte der Situation vom Wissensinhalt zu trennen. Dieser Prozess wird als Entkontextualisieren bezeichnet. Dadurch verliert der Lernende die Assoziation einer Aufgabe mit einem bestimmten Kontext und allmählich tritt das zugrunde liegende Prinzip hervor Perkins und Salomon (1989). Entkontextualisieren von Wissen ist jedoch nicht ausreichend, zusätzlich müssen Lernende lernen, wann und wo welches Wissen angewendet werden muss Wiggins (1993). Diese Erkenntnis deckt sich mit den Ergebnissen von Gick und Holyoak (1980), bei denen die Teilnehmer einen höheren Transfer aufwiesen, wenn auf die Ähnlichkeit der Situationen hingewiesen wurden.

## Problemorientierter Unterricht

Williams (1992) untersuchte viele Lernsituationen im Medizinischen Studium auf ihre Möglichkeiten zu Transfer. Sie stellt fest, dass das Wissen, welches in Vorlesungen gelernt wurde, im klinischen Teil der Ausbildung vergessen ist. Ein Grund dafür sind, dass in vielen Lehrbüchern und Vorlesungen theoretisches Wissen losgelöst von Anwendungen dargestellt werden. So werden Fragen beantwortet, welche sich Lernende nicht stellen und daher von diesen nicht auf konkrete Problemsituationen angewendet werden können. Diese Erkenntnis gilt nicht nur für Mediziner, sondern wurde auch in anderen Fachdisziplinen nachgewiesen. So bedauert Shuell (1996), dass zukünftige Lehrpersonen faktisches Wissen lernen, anstelle von anwendungsbezogenem Wissen. So lernen sie etwas *über* das Unterrichten jedoch nichts darüber, *wie* zu unterrichten ist.

Um diese Probleme zu vermeiden wurde problemorientierte Unterrichtsgelegenheiten entwickelt und untersucht (siehe unter anderem Barrows (1985), Michael et al. (1993), Shuell (1996), Corte (2003), Reusser (2005), Fässler (2007) und Pea (2013)). Das Ziel dabei ist, dass Wissen in möglichst lebensnahen Kontexten zu erwerben. Dies führt dazu, dass bei der Anwendung der Kontext ähnlich zu dem Kontext ist, unter welchem das Wissen erworben wurde.

### 2.1.6 Zusammenfassung zu Transfer

Es wurde in dem letzten Abschnitt versucht einen Überblick über der Begriff des Transfers zu geben und zu zeigen wie der Begriff des Kontextes damit verknüpft ist. Zuerst wurde eine historische Übersicht, über die wichtigsten Untersuchungen zu Transfer gegeben, um den Wandel des Begriffes des Transfers aufzuzeigen. Als Vorbereitung für den nächsten Abschnitt wurde noch der Begriff des Transfers elementarisiert. Darauf Aufbauend wurden die Konsequenzen für den Unterricht, welcher transferierbares Wissen fördern soll zusammengefasst. Im nächsten Abschnitt geht es um den Begriff der Kompetenz und deren Verknüpfung mit dem Begriff des Transfers und Kontext.

---

## 2.2 Kompetenz

Nachdem ein Überblick über den Transfer erarbeitet wurde, soll in diesem Abschnitt versucht werden die Konsequenzen aus der Betrachtung zu Transfer mit dem Kompetenzbegriff zu verknüpfen.

### 2.2.1 Bildungsreformen

In den letzten Jahrzehnten fand international ein Wandel in der Bildungspolitik statt. In der Vergangenheit wurde der Fokus auf den Input des Bildungssystems gelegt. In den letzten Jahren fand eine Erweiterung der Perspektive statt und auch der Output des Bildungssystems wurde beachtet. Das Ziel dabei ist, die Qualität des Bildungssystems fassbar zu machen und soll helfen die Ressourcen effektiv einzusetzen.

Dieser Perspektivwechsel wurde von den grossen Bildungsstudien (PISA (PISA-Konsortium Deutschland 2004), TIMSS (Martin und Mullis 2003) und IGLU (Bos et al. 2003)) in den letzten Jahren ausgelöst. Diese führten zu einem Wandel, sowohl in der Forschung als auch in der politischen Diskussion über das Bildungssystem. So wurden die Resultate des Bildungssystems in den Vordergrund gerückt. Insbesondere die Definition von Standards und deren Verankerung im gesamten Bildungssystem sind neu. Davor wurden Standards meistens durch strukturelle Vorgaben umgesetzt (Lehrpläne, Stundentafeln und Schulorganisation). Diese Vorgaben haben einen Einfluss auf den Input des Bildungssystems. Die Qualität des Bildungssystems wurde jedoch nur sehr gering über die Überprüfung der erreichten Ergebnisse (Leistung der Schülerinnen und Schüler, Übertrittsquoten und Abschlussprüfungen) überprüft. Es wurde implizit angenommen, dass der Input einen Einfluss auf das Ergebnis des Bildungssystems als ganzes hat.

Neu ist, dass die Steuerung des Bildungssystems vermehrt über den Output erfolgen soll. So soll die Leistung des Bildungssystems messbar gemacht werden und objektiv vergleichbar. Mit den bisherigen Leistungserhebungen auf Klassen oder Schulstufe, lässt sich der Output des Schulsystems nicht akkurat beschreiben, da festgelegte Messstandards fehlten. So wurden im Zuge der Entwicklung von Bildungsstandards kompetenzbezogene Niveaus eingeführt, welche einen Aufschluss über die erreichten Kompetenzen eines Schülers oder Schülerin geben soll (Oelkers et al. 2008).

Diese Bemühungen führten in vielen Ländern zur Entwicklung von neuen Bildungsstandards (Berner und Stolz 2006). In der Schweiz wurde dies von der EDK Schweizer Konferenz der Kantonalen Erziehungsdirektoren (2004) unter dem Title „Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat)“ angestossen. In Deutschland wurde neue Bildungsstandards von der Kultusministerkonferenz (2004) verabschiedet.

Im englischsprachigen Raum fanden diese Diskussionen bereits früher statt. So wurde in Neuseeland bereits zu Beginn der 1990er Jahren ein „Outcome-based“ Curriculum verabschiedet (McGee 1996). Auch in Australien wurde ein ähnliches Bildungskonzept 2000, unter dem Name „outcomes-based education (OBE)“, umgesetzt (Killen 2000). Auch in England wurde zu Beginn des Jahrtausends Bildungsreformen gefordert (Millar und Osborne 1999), welche dann um 2005 umgesetzt wurden (Huber et al. 2006).

---

### 2.2.2 Definition von Kompetenz

Der Begriff der Kompetenz wird im Moment sowohl fachlich als auch politisch sehr stark diskutiert. So spricht Weinert (2001) von einer Inflation des Kompetenzbegriffes. Der Begriff der Kompetenz, welcher in den Bildungsstandards (sowohl der Schweiz als auch von Deutschland) verwendet wird, basiert auf der Arbeit von Klieme (2004). Klieme (2004) unterscheidet verschiedene Varianten des Kompetenzbegriffes:

1. Kompetenz als kognitive Leistungsdisposition, welche es Personen erlaubt unterschiedliche Aufgaben zu lösen.
2. Kompetenz als kontextspezifische kognitive Leistungsdisposition, welche sich auf spezifische Kontexte bezieht. Dieser Kompetenzbegriff wird oft mit Kenntnisse, Routinen oder Fertigkeiten bezeichnet.
3. Kompetenz als motivationaler Orientierungen, welche notwendig ist um eine Aufgabe zu bewältigen.
4. Handlungskompetenz, als Integration der vor-gängigen Kompetenzbegriffe, im Bezug auf die Anforderungen eines genau definierten Handlungskontextes.
5. Metakompetenzen als Strategiewissen oder Motivation, welche die Anwendung und den Erwerb anderer Kompetenzen erleichtert.
6. Schlüsselkompetenzen als generalisierbare kontextspezifische kognitive Leistungsdispositionen. Das heisst Kompetenzen, welche auf viele verschiedene Situationen angewendet werden können, wie zum Beispiel mathematische oder sprachliche Kenntnisse.

#### Abgrenzung von Kompetenz und Intelligenz

Problematisch an der Definition des Kompetenzbegriffes ist, dass der Kompetenzbegriff schwierig von der Definition der allgemeinen Intelligenz zu unterscheiden ist, insbesondere der erste Punkt. Weinert (2001) empfiehlt daher eine Einschränkung des Kompetenzbegriffes. So sollen Kompetenzen auf einen eingeschränkten Raum von Kontexten und Situationen bezogen werden und allgemeine intellektuellen Fähigkeiten ausgeschlossen werden. Dies begründet Weinert (2001) damit, dass allgemeine intellektuelle Fähigkeiten eine Grundausstattung des Menschen sind und nicht erworben werden können und daher nur sehr begrenzt trainiert werden können. Zusätzlich schränkt er den Kompetenzbegriff weiter ein, indem er affektive und motivationale Aspekte nicht einbezieht. Der Begriff der Kompetenz soll daher auf spezifische Kenntnisse angewendet werden, welche notwendig sind, um genau definierte Ziele zu erreichen. Dies führt zu einer Abgrenzung zum Intelligenzkonzept, da damit Fähigkeiten assoziiert werden, welche ohne spezifisches Vorwissen auf neue Problemstellungen angewendet werden sollen. Daher ist der Begriff der Kompetenz stärker mit spezifischen Kontexten verbunden, während die Intelligenz sich generalisieren lässt. Dies führt jedoch zu einem weiteren Problem, da bei breiteren Kontexten die Abgrenzung zwischen Kompetenz und Intelligenz schwieriger wird (Hartig und Klieme 2006).

Ein weiterer Unterschied zwischen dem Kompetenz- und dem Intelligenzkonzept beruht auf der Lernbarkeit. So fordert Baumert, Stanat und Demmrich (2001, S. 22), dass Kompetenzen „prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse und Strategien“ sind. Dies bedeutet, dass Kompetenzen durch schulischen Unterricht gefördert

und erweitert werden können. Daher sind dies Leistungen, welche durch den Schulbesuch verbessert werden sollten und sind daher für das Bildungsmonitoring von Interesse. Intelligenz wird hingegen als relativ stabil betrachtet, da Intelligenz hauptsächlich von genetischen Faktoren abhängt (Shakeshaft et al. 2013). Daher sollte theoretisch der Schulbesuch keine direkte Verbesserung der Intelligenzleistung zur Folge haben. Dies führt auch zu einem weiteren Unterschied zwischen der Intelligenz und der Kompetenz. So kann die Kompetenz in einem bestimmten Bereich bei null liegen, da die Erfahrungen, welche zu dem Erwerb der Kompetenz führen noch nicht gemacht wurden. Bei der Intelligenzleistung ist dies nicht möglich, da jeder Mensch sich diese Grundfertigkeiten irgendwo angeeignet haben sollte.

Des Weiteren gibt es bei Erstellung von Leistungsmessungen Unterschiede. Bei Intelligenztests werden bestimmte Primärfaktoren verwendet, z.B. dreidimensionales Denken, Gedächtnis, usw., welche Unterschiede zwischen einzelnen Personen aufzeigen sollen. Kompetenzen hingegen werden durch die Anforderungen definiert (Rychen und Salganik 2001). In anderen Worten, Kompetenzen werden durch die relevanten Aufgaben definiert, welche von den Untersuchten gelöst werden sollen. So werden in HarmoS Kompetenzen in einem dreidimensionalen Modell definiert, Themengebiete, Kompetenzaspekt und Kompetenzniveau (Konsortium HarmoS Naturwissenschaften+ 2010). Die Kompetenzaspekte werden nicht wie bei der Intelligenz über psychischen Prozesse definiert, sondern aus spezifischen Anforderungen in spezifischen Kontexten abgeleitet. Diese Unterschiede führen dann auch zu einer unterschiedlichen Konstruktion von Leistungstests. Intelligenztests sollten so konstruiert sein, dass möglichst wenig Vorwissen für das Lösen von Aufgaben notwendig ist. Tests, welche auf Kompetenzen abzielen, wie z.B. PISA, wurden mit dem Ziel entwickelt, Aufgaben in realitätsnahen Kontexten zu stellen.

Trotz all dieser Unterschiede werden typischerweise hohe Korrelation zwischen Intelligenz- und Kompetenzleistungen festgestellt. So fand Rindermann (2006) meistens eine sehr hohe Korrelation ( $>0.7$ ) zwischen Kompetenzen und anderen Massen für kognitive Fähigkeiten. Es ist aber mit den Resultaten von PISA nicht möglich festzustellen ob dies daran liegt, dass Schüler und Schülerinnen eine höhere Kompetenz haben, weil sie intelligent sind, oder ob sowohl Intelligenz als auch Kompetenz durch die Schulbildung geprägt werden (Hartig und Klieme 2006).

### 2.2.3 Kompetenz und Transfer

Der Kompetenzbegriff wird in der Literatur sehr unterschiedlich definiert (Klieme 2004; Weinert 2001). Dennoch ist die Definition des Kompetenzbegriffes für internationale Studien wie PISA (PISA-Konsortium Deutschland 2004), TIMSS (Martin und Mullis 2003) und IGLU (Bos et al. 2003) elementar.

Interessant ist, dass trotz der Einschränkung des Kompetenzbegriffes auf spezifische Kontexte, immer noch davon ausgegangen wird, dass die Kompetenz generalisierbar ist und teilweise auf andere Situationen übertragen werden kann Hartig und Klieme (2006).

Im Abschnitt 2.1.5 wurde herausgearbeitet, welche Eigenschaften von Unterricht zu besserer Transferleistung führen kann. Auch Lersch (2007) gibt Vorschläge wie kompetenzfördernder Unterricht gestaltet sein sollte. So fordert er, dass der Unterricht „viel stärker von den erfor-

derlichen Lernprozessen und -gelegenheiten her konzipiert werden müsste und eben nicht nur von einer kontinuierlichen Abfolge von Inhalten“. Dies deckt sich mit der Forderung von Mietzel (2007) für das Entkontextualisieren von Unterricht um Transferleistung zu fördern. Auch die Problemorientierung von Lerngelegenheiten wird von Lersch (2007) für kompetenzfördernden Unterricht als wichtig gehalten, insbesondere fordert er, dass „systematische Wissensvermittlung [...] um variable Anwendungssituationen“ ergänzt werden sollten. Zusätzlich fordert er, dass realistische Lernsituationen angeboten werden sollten, in anderen Worten: die Lerngelegenheiten sollten mit dem Ziel der Kompetenz übereinstimmen, da der Erwerb der Kompetenz ja kontextspezifische erfolgt (Klieme 2004).

## 2.3 Kompetenz des skalenbasiertes Messens

Im Rahmen von ExKoNawi hands-on Experimentiertests wurde ein Modell entwickelt um verschiedene hands-on Kompetenzen von Schülerinnen und Schüler auf der Sekundarstufe I in der Schweiz zu messen werden (Metzger et al. 2013). Einer der Kompetenzen welche mit ExKoNawi hands-on Experimentiertests gemessen werden soll, ist die Kompetenz des *skalenbasiertes Messen* (Gut et al. 2014). Die Definition dieser Kompetenz basiert auf der Arbeit von Munier, Merle und Brehelin (2013). In dieser Kompetenz geht nach (Gut et al. 2014) darum „quantitative Grössen mit gegebenen Messinstrument genau [zu] messen“. Bei dieser Kompetenz gibt es drei Teilbereiche die eine wichtige Rolle spielen. Zum einen müssen die Schülerinnen und Schüler entscheiden, welches Messinstrument besser für eine Messung geeignet ist. Ein weiterer Teilaspekt ist, dass sie die Messung mehrmals wiederholen um eine genauere Abschätzung des Resultates bekommen. Zusätzlich zu diesen Aspekten müssen sie auch das Messinstrument korrekt verwenden (Munier, Merle und Brehelin 2013; Gut et al. 2014).

Ein wichtiger Aspekt der Kompetenz des skalenbasierten Messens ist, dass diese Kompetenz ohne einen inhaltlichen oder fachlichen Kontext definiert wurde. Was bedeuten sollte, dass das Kompetenzniveau, welches ein Schüler oder eine Schülerin erreichen könnte unabhängig des fachlichen oder inhaltlichen Kontextes sein müsste, in welcher die Messung der Kompetenz stattfinden sollte.

## 2.4 Forschungsfrage

Dieses Modell führt nun daher zur Frage, ist das erreichbare Kompetenzniveau von Schülerinnen und Schüler tatsächlich unabhängig des inhaltlichen und fachlichen Kontextes? Insbesondere auch aus dem Aspekt, dass sich kontextspezifische Kompetenzen und Transferleistungen grundsätzlich nicht gegenseitig ausschliessen. Guter kompetenzorientierter Unterricht unterstützt hingegen sogar die Fähigkeiten das Wissen zu transferieren. Diese Erkenntnis führt nun jedoch zu der Frage:

Ist eine gewisse Kompetenz (hier skalenbasiertes Messen) von Lernenden auf der Sekundarstufe I in unterschiedlichen Kontexten gleich verfügbar?

Diese Frage verknüpft sehr stark den Begriff des Transfers mit dem Kompetenzbegriff. Im Bezug auf den Transferbegriff müssen Lernende eine Transferleistung erbringen, da sie

diese gewisse Kompetenz auf verschiedene Kontexte anwenden müssen. Um diese Frage zu beantworten soll in der vorliegenden Arbeit untersucht werden ob die erreichten Kompetenzniveaus des skalenbasierten Messens unabhängig des fachlichen oder inhaltlichen Kontextes sind.



## 3 Untersuchungsanlage

### 3.1 Anforderungen

Um die vorliegende Fragestellung zu beantworten, ist es notwendig mehrere Test zu verwenden, welche die Kompetenz des skalenbasierten Messens messen. Zusätzlich müssen die Tests die Kompetenz des skalenbasierten Messens unter verschiedenen Kontexten messen. Es wurden zwei existierende Test aus dem ExKoNawi Projekt verwendet. Der eine war aus dem Fachbereich Chemie, bei welchem eine Temperatur gemessen werden musste. Der zweite Test war aus dem Fachbereich Physik, bei dem eine Kraft gemessen wurde. Zusätzlich wurde ein dritter Test neu entwickelt, bei welchem eine Temperatur Messung im Fach Physik durchgeführt wurde. Die Testerstellung wird genauer in Sichau (2015) detailliert beschrieben. Der dritte Test wurde so entworfen, damit einmal der inhaltliche Kontext verändert werden kann (Kraftmessung versus Temperaturmessung), bei gleichem fachlichen Kontext und zum anderen der fachliche Kontext verändert werden kann ohne den inhaltlichen Kontext zu verändern.

### 3.2 Umsetzung

Die Tests wurden zusammen mit einem Fragebogen an vier Klassen der Sek 1 A durchgeführt. In jeder Klasse wurden vier Gruppen gebildet, welche die Tests in unterschiedlicher Reihenfolge durchführten. Dafür gab es zwei Gründe. Zum einen war nur Material für 11 Tests verfügbar. Daher konnten die Tests nicht in voller Klassenstärke durchgeführt werden. Dies führte zur Bildung von zwei Gruppen, bei welcher eine zuerst den Fragebogen ausfüllte und die andere Gruppe den Fragebogen am Ende ausfüllte. Zusätzlich wurde noch der zweite und dritte Test in jeder Gruppe vertauscht um zu untersuchen ob Müdigkeit oder die Wiederholungen Einfluss auf die Test-Ergebnisse haben. Die Tabelle 3.1 gibt eine Übersicht über die Gruppeneinteilung der Schülerinnen und Schüler innerhalb einer Klasse an.

Gruppe FABC	Gruppe FACB	Gruppe ABCF	Gruppe ACBF
Fragebogen (F)	Fragebogen (F)	Temperatur Physik 305 (A)	Temperatur Physik 305 (A)
Temperatur Physik 305 (A)	Temperatur Physik 305 (A)	Kraft Physik (B)	Temperatur Chemie (C)
Kraft Physik 301 (B)	Temperatur Chemie 201 (C)	Temperatur Chemie 201 (C)	Kraft Physik 301 (B)
Temperatur Chemie 201 (C)	Kraft Physik 301 (B)	Fragebogen (F)	Fragebogen (F)

Tabelle 3.1: Aufteilung der Gruppen, innerhalb einer Klasse



Die Namen der Gruppen aus Tabelle 3.1 wurden auch für die Kodierung der Tests verwendet, sodass jeder Test einer Gruppe zuordenbar ist.

Die vier Klassen waren alle von derselben Schulstufe (7. Schuljahr) jedoch in verschiedenen Gemeinden. Die Klassen in Glattbrugg hatten beide dieselbe Lehrperson, alle anderen Klassen hatten eine unterschiedliche Lehrpersonen. Einen Überblick über die wichtigsten Daten zu den einzelnen Klassen befindet sich in Tabelle 3.2.

	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Ort	Glattbrugg	Glattbrugg	Stadt Zürich	Stadt Schaffhausen
Anzahl SuS	15	13 (+1 nur einen Test)	22	22
Datum der Durchführung	6.11.14	6.11.14	12.11.14	11.12.14
Uhrzeit	8:20-10:00	10:20-12:00	10:20-12:05	13:15-14:45
Versuchsleiter	Pitt Hild und David Sichau	Pitt Hild und David Sichau	Pitt Hild und David Sichau	Martina Minges und David Sichau

Tabelle 3.2: Aufteilung der Gruppen, innerhalb einer Klasse

Alle Klassen wurden für die Durchführung in zwei Gruppen aufgeteilt. Zum einen konnten so die Schülerinnen und Schüler mit mehr Abstand positioniert werden um die Ablenkung zu reduzieren. Andererseits konnten so die Schülerinnen und Schüler, welcher der Videoaufnahme nicht zugestimmt wurden in ein Zimmer gesetzt werden, wo keine Videoaufnahme stattgefunden hat. Die Erlaubnis zur Videoaufnahme wurde im Vorhinein zur Durchführung von den Klassenpersonen organisiert und eingesammelt.

## 3.3 Durchführung

### 3.3.1 Vorbereitung

Für die Durchführung in den einzelnen Klassen wurden alle Tests in Boxen vorbereitet, sodass zwischen den Tests nur die Boxen ausgetauscht werden mussten. In jeder Box waren alle Materialien, welche für die Durchführung des Versuches notwendig waren vorbereitet, sodass die Schülerinnen und Schüler alle notwendigen Materialien in dieser Box finden konnten.

Zusätzlich wurden die Auswertungsbögen in der richtigen Reihenfolge und bereits mit einer Kodierung versehen, in einem Schnellhefter bereitgestellt. Ein für die Durchführung vorbereiteter Klassenraum ist im Bild 3.1 ersichtlich.

Im Bild 3.1 sieht man auch gut, wie die Kamera für die Videoaufwertung aufgestellt wurde. Die Videoaufnahme wurde bevor die Schülerinnen und Schüler den Klassenraum betreten haben gestartet, um die Ablenkung durch die Kamera möglichst gering zu halten.



Abbildung 3.1: Klassenzimmer für die Durchführung des ersten Durchganges vorbereitet.

### 3.3.2 Durchführung

Nachdem die Schülerinnen und Schüler in die beiden Räume aufgeteilt wurden, wurden Sie von den Versuchsleitern jeweils begrüßt. Die Begrüssung war stichwortartig vorbereitet, damit alle Klassen die gleichen Informationen erhielten und durch die Begrüssung die Testergebnisse nicht beeinflusst werden. Dabei wurde darauf hingewiesen, dass die Experimente keine Leistungskontrolle darstellt und alle Ergebnisse anonymisiert sind. Es wurde auch ein grober Überblick über den Ablauf gegeben. Im Raum in dem eine Videoaufnahme gemacht wurde, wurden die Schülerinnen und Schüler darüber informiert.

Nach der Begrüssung wurden die Schülerinnen und Schüler aufgefordert mit den Tests anzufangen. Während der Zeit in welcher die Tests durchgeführt wurden, gaben die Versuchsleiter jeweils kurze Zeit Informationen und forderten die Schülerinnen und Schüler auf ihre Ergebnisse zu verschriftlichen.

Nach dem ersten Test (nach 20 Minuten) wurde eine Pause von fünf Minuten durchgeführt, in dieser wurden die Boxen ausgetauscht, sodass alle Schülerinnen und Schüler den nächsten hands-on Experimentiertest vor sich hatten. Die Schülerinnen und Schüler wurden aufgefordert sich innerhalb des Klassenraumes zu bewegen. Nach dem zweiten Test wurde eine grosse Pause durchgeführt, in welcher die Schülerinnen und Schüler das Schulzimmer verlassen hatten. Nach dem dritten Test wurde wieder eine kurze fünf-minütige Pause durchgeführt. Während den Test wurden den Schülerinnen und Schülern nur Fragen bei Unklarheiten beantwortet, inhaltliche Fragen oder Fragen zum korrekten Vorgehen wurden nicht beantwortet.

### 3.3.3 Nachbereitung

Nachdem die Tests durchgeführt wurden, wurden die Auswertungsbögen eingesammelt und von David Sichau erst kodiert. Es wurde eine Zweitkodierung vor 15 % der Auswertungsbögen von Pitt Hild durchgeführt. Die 11 Auswertungsbögen zur Zweit-Kodierung wurden per zufällig (random generator) ausgewählt, um sicherzugehen das ein Bias ausge-

geschlossen werden kann. Insgesamt wurden 72 Auswertungsbögen vollständig ausgefüllt.

Die Videoaufnahmen wurden geschnitten, sodass nur noch die einzelnen Tests sichtbar sind. Dies wurde gemacht um zu vermeiden, dass Aktionen der Schülerinnen und Schüler in der Pause einen Einfluss auf die Bewertung in der Test Situation haben. Insgesamt ist Material zu 8 Schülerinnen und Schüler verwertbar, da die andern zu weit entfernt sind und daher ihre Aktionen nicht beobachtbar waren.

## 4 Ergebnisse

### 4.1 Kodierung

Wie bereits geschrieben wurde die Erstkodierung von David Sichau durchgeführt. Es wurden eine Zweitkodierung von 15 % zufällig ausgewählten (per Random Generator) Auswertungsbögen von Pitt Hild durchgeführt. Es wurden dabei die gleichen Kodierschemata verwendet, welche sich im Anhang der Arbeit befinden.

#### 4.1.1 Items

Es gab insgesamt elf Items welche nach dem Kodierschemata kodiert wurden.

Die Items wurden auf Interrater-Reliabilität untersucht. Dafür wurden die prozedurale Übereinstimmung  $p_0$  und zusätzlich noch das ungewichtete Cohen's Kappa  $\kappa$  als Zufalls-korrigierter Koeffizient berechnet. Bei einem Teil der Datensätze war dies mathematisch nicht möglich (Division durch 0), daher können nicht für alle Items ein Cohen's Kappa angegeben werden. In Tabelle 4.1 sind alle Ergebnisse zusammengefasst.

Code erhältlich auf:

**GitHub**

<http://git.io/mk9z-Q>

Item	201		301		301	
	$p_0$	$\kappa$	$p_0$	$\kappa$	$p_0$	$\kappa$
1.1	1	1	0.91	0.74	0.91	0.79
1.2	0.91	0.81	1	/	1	1
2.1	0.81	0.67	0.81	0.74	1	1
3.1	1	1	0.91	0.81	1	1
3.2	1	/	1	1	0.91	0.82
4.1	0.91	0.79	0.81	0.65	0.91	0.81
4.2	0.91	0.62	0.91	0.79	0.91	0.74
4.3	1	/	1	/	1	/
4.4	1	/	1	/	1	/
5.1	1	/	1	/	1	/
5.2	0.91	/	1	1	0.91	0.78

Tabelle 4.1: Übereinstimmung der Kodierungen für die einzelnen Items ( $p_0$ ) und Cohens Kappa  $\kappa$ . Für die drei Tests 201 (Chemie-Temperatur), 301 (Physik Kraft) und 305 (Physik Temperatur)

### 4.1.2 Qualitätsstandards

Aus den elf Items wurden fünf Qualitätsstandards entwickelt (**Hild2014a** ).

#### Qualitätsstandard 1

Im Qualitätsstandard 1 geht es um das korrekte und präzise messen. Dieser Qualitätsstandard wird nur erreicht wenn Item 1.1 (Richtige Tendenz des Resultates) und Item 1.2 (Ist das Resultat vollständig und korrekt) zusammen mindestens 1 ergeben.

#### Qualitätsstandard 2

Bei Qualitätsstandard 2 wird die Dokumentation der Messung bewertet . Dieser Qualitätsstandard wird nur erreicht wenn Item 1.2 (Werden alle Messungen und Messergebnisse vollständig dargestellt) mindestens den Wert von 2 erreicht hat.

#### Qualitätsstandard 3

Im dritten Qualitätsstandard wird das Begründen des richtigen Messinstrumentes bewertet. Dieser Standard wird nur erreicht wenn Item 3.1 (Ist das Korrekte Messinstrument gewählt worden) und Item 3.2 (Wird die Wahl des Messinstrumentes korrekt begründet) zusammen zwei ergeben.

#### Qualitätsstandard 4

Qualitätsstandard 4 beurteilt die Messwiederholung. Es wird aus Item 4.1 (mehrmaliges Messen), 4.2 (identische Messung), 4.3 (wurde Mittelwert gebildet) und 4.4 (korrekter Mittelwert) gebildet. Diese Level wird erreicht wenn die Items addiert mindestens zwei ergeben.

#### Qualitätsstandard 5

Der letzte Qualitätsstandard 5 zeigt auf, inwiefern die Schülerinnen und Schüler Fehlerquellen der Messung begründen können. Dieser Standard besteht aus Item 5.1 (Fehlerkategorien nennen) und 5.2 (Verbesserungsvorschläge) welche zusammen mehr als eins ergeben müssen.

#### Erreichte Qualitätsstandards

In Tabelle 4.2 wird ein Überblick über die erreichten Qualitätsstandards aller Schülerinnen und Schüler gegeben. Zusätzlich werden auch die bedingten Qualitätsstandards angegeben, welche nur erreicht werden können, wenn der vorhergehende Qualitätsstandard erreicht wurde.

Test	$p_{Q1}$	$p_{QS1}$	$p_{Q2}$	$p_{QS2}$	$p_{Q3}$	$p_{QS3}$	$p_{Q4}$	$p_{QS4}$	$p_{Q5}$	$p_{QS5}$
201	0.51	0.51	0.34	0.27	0.05	0.04	0.08	0.03	0.16	0.03
301	0.62	0.62	0.31	0.31	0.09	0.04	0.09	0.01	0.39	0.01
305	0.72	0.72	0.30	0.29	0.35	0.14	0.11	0.01	0.50	0.01

Tabelle 4.2: Zusammenfassung der erreichten Qualitätsstandards, wobei  $p_{Q1} - p_{Q5}$  den unbedingten Qualitätsstandards entsprechen. Die bedingten Qualitätsstandards werden mit  $p_{QS1} - p_{QS5}$  bezeichnet.

### 4.1.3 Niveau

Basierend auf den Qualitätsstandards wurden zwei Niveaus gebildet, welche das erreichte Niveau der Schülerinnen und Schüler bei der Kompetenz des skalenbasierten Messens bezeichnen. Die Niveaus können einen Wert zwischen 0 und 5 annehmen. Eine Übersicht über die erreichten Niveaus wird in Tabelle 4.3 gegeben.



Test	uLev						cLev					
	0	1	2	3	4	5	0	1	2	3	4	5
201	0.36	0.24	0.22	0.13	0.03	0.03	0.40	0.24	0.32	0.01	0	0.03
301	0.31	0.21	0.29	0.14	0.03	0.03	0.42	0.28	0.26	0.01	0	0.03
305	0.13	0.19	0.24	0.31	0.11	0.03	0.22	0.43	0.18	0.13	0.01	0.03

Tabelle 4.3: Prozedural erreichte Niveaus aller Schülerinnen und Schüler. Das unbedingte Niveau wird mit uLev und das bedingte Niveau mit cLev bezeichnet.

#### Unbedingtes Niveau

Dieses Niveau ist der Summenscore der einzelnen unbedingten Qualitätsstandards. In der Arbeit wird dieses Level mit *uLev* abgekürzt.

#### Bedingtes Niveau

Dieses Niveau ist der Summenscore der bedingten Qualitätsstandards. Dieses Niveau wird mit *cLev* abgekürzt.

## 4.2 Fragebogen

Im standardisierten Teil des Fragebogens wurden Fragen zum absoluten Selbstkonzept nach SESSKO gestellt (Schöne et al. 2002). Die verwendeten Fragen sind in Tabelle 4.4 aufgeführt.

Skala	Frage	$\alpha_d$
SESSKO 18(a)	Ich bin für die Schule sehr begabt.	0.71
SESSKO 19(a)	Neues zu lernen fällt mir schwer.	0.76
SESSKO 20(a)	Ich bin sehr intelligent.	0.71
SESSKO 21(a)	Ich kann in der Schule viel.	0.72
SESSKO 22(a)	In der Schule fallen mir viele Aufgaben schwer.	0.74

Tabelle 4.4: Fragen von SESSKO zur Skala „Schulisches Selbstkonzept - absolut“ (Schöne et al. 2002).  $\alpha_d$  bezeichnete das standardisierte Cronbach Alpha wenn dieses Item weggelassen würde.

Zusätzlich wurden nach Dierks, Höffler und Parchmann (2014) Fragen zum Selbstkonzept zu Schulversuchen entwickelt und angepasst. Die entwickelten Fragen sind in Tabelle 4.5 aufgezeigt.

Kürzel	Frage	$\alpha_d$
NatSK1	Schulversuche liegen mir nicht besonders.	0.65
NatSK2	Schulversuche würde ich viel lieber machen, wenn sie nicht so schwer wären.	0.69
NatSK3	Schulversuche fallen mir schwerer als vielen meiner Mitschüler/innen.	0.65
NatSK4	Bei manchen Schulversuche weiss ich gleich: „Das verstehe ich nie.“	0.65
NatSK5	Für Schulversuche habe ich einfach keine Begabung.	0.63
NatSK6	Mit den Aufgaben bei Schulversuche komme ich besser zurecht als viele meiner Mitschüler/innen	0.67
NatSK7	Ich denke, ich bin für Schulversuche begabter als viele meiner Mitschüler/innen.	0.66

Tabelle 4.5: Fragen zum Selbstkonzept bei Schulversuchen abgewandelt nach Dierks, Höffler und Parchmann (2014).  $\alpha_d$  bezeichnete das standardisierte Cronbach Alpha wenn dieses Item weggelassen würde.

Es wurde die innere Konsistenz beider Skala überprüft. Bei den der Skala „Schulisches Selbstkonzept - absolut“ wurde ein standardisiertes Cronbach  $\alpha$  von 0.77 erreicht. Die Anzahl vollständig ausgefüllter Fragebögen betrug dabei 69. Alle unvollständigen Items wurden vor der Analyse entfernt. Bei der Skala zum Selbstkonzept bei Schulversuchen wurde ein standardisiertes Cronbach  $\alpha$  von 0.69 erreicht. Insgesamt konnten dabei 64 vollständige Fragebögen ausgefüllt werden.

Code erhältlich auf:

**GitHub**

<http://git.io/WyJH6Q>

## 4.3 Unterschiede zwischen den Klassen

Um festzustellen, ob alle Datensätze der einzelnen Klassen kombiniert werden dürfen wurden zuerst alle Klassen einzeln gegeneinander auf folgende Nullhypothese überprüft: Besteht kein Unterschied in den Qualitätsstandards zwischen den einzelnen Klassen?

Aufgrund der geringen Anzahl an Datenätzen in den einzelnen Klassen wurde der exakter Test nach Fisher verwendet. Dieser Test wurde für jeden Qualitätsstandard (bedingt und unbedingt) einzeln jeder Klasse gegen jede andere Klasse durchgeführt. Bei keinem der 60 Tests konnte die Nullhypothese abgelehnt werden ( $p \geq 0.05$ ). Daher gibt es keinen signifikanten Unterschied zwischen den erreichten Qualitätsstandards und den einzelnen Klassen.

Code erhältlich auf:

**GitHub**

<http://git.io/0DOelQ>



## 5 Diskussion

## 6 Ausblick

# Literaturverzeichnis

- Anderson, John R., Lynne Reder und Herbert A. Simon. 1996. "Situated Learning and Education". *Educational Researcher* 25, Nr. 4 (Mai): 5–11.
- Barrows, Howard S. 1985. *How to design a problem-based curriculum for the preclinical years*.
- Baumert, Jürgen, Petra Stanat und Anke Demmrich. 2001. "PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie". In *PISA 2000*, herausgegeben von Manfred Weiß. Opladen.
- Berner, Esther, und Stefanie Stolz. 2006. *Literaturanalyse zu Entwicklung, Anwendung und insbesondere Implementation von Standards in Schulsystemen: Nordamerika*. Technischer Bericht. Zürich: Pädagogisches Institut der Universität Zürich.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther. 2003. *Erste Ergebnisse aus IGLU*. April 2003.
- Brophy, Jere. 1992. "Probing the subtleties of subject-matter teaching." *Educational Leadership*.
- Claxton, Guy. 1990. *Teaching to Learn: A Direction for Education*. Cassell education. Cassell.
- Corte, Erik De. 2003. "Designing learning environments that foster the productive use of acquired knowledge and skills". Kap. 2 in *Powerful Learning Environments: Unraveling Basic Components and Dimensions*, 21–33.
- Detterman, Douglas K. 1993. "The case for the prosecution: Transfer as an epiphenomenon". Kap. 1 in *Transfer on trial: Intelligence, cognition, and instruction*, herausgegeben von Douglas K (ED) Detterman und Robert J (ED) Sternberg, 1–24. Ablex Publishing.
- Dierks, Pay Ove, Tim Höffler und Ilka Parchmann. 2014. "Interesse von Jugendlichen an Naturwissenschaften Ist es wirklich so schlecht wie sein Ruf?" *CHEMKON* 21 (3): 111–116.
- Duncker, Karl, und Lynne S. Lees. 1945. "On problem-solving." *Psychological monographs* 58 (5): i.
- EDK Schweizer Konferenz der Kantonalen Erziehungsdirektoren. 2004. *HARMOS Zielsetzungen und Konzeption Juni 2004*. Technischer Bericht. Bern: EDK Schweizer Konferenz der Kantonalen Erziehungsdirektoren.
- Fässler, Lukas. 2007. "Das 4-Schritte-Modell". Diss.
- Ferguson, George A. 1956. "On transfer and the abilities of man." *Canadian journal of psychology* 10, Nr. 3 (September): 121–31.

- Gick, Mary L., und Keith J. Holyoak. 1980. "Analogical problem solving". *Cognitive psychology* 355:306–355.
- Godden, D.R., und A.D. Baddeley. 1975. "Context-dependent memory in two natural environments: On land and underwater". *British Journal of psychology*.
- Greeno, James G., Allan M. Collins und Lauren B. Resnick. 1996. "Cognition and learning". In *Handbook of Educational Psychology*, herausgegeben von D. Berliner und R. Calfee, 14–46. New York.
- Gut, Christoph, Pitt Hild, Susanne Metzger und Josiane Tardent. 2014. "Projekt ExKoNa-wi: Modell für hands-on Assessments experimenteller Kompetenzen". In *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht*, herausgegeben von Sascha Bernholt, 171–173.
- Hartig, Johannes, und Eckhard Klieme. 2006. "Kompetenz und Kompetenzdiagnostik". Kap. 3 in *Leistung und Leistungsdiagnostik*, herausgegeben von Karl Schweizer, 127–143. Springer Berlin Heidelberg.
- Huber, Christina, Martina Späni, Claudia Schmellentin und Lucien Criblez. 2006. *Bildungsstandards in Deutschland, Österreich, England, Australien, Neuseeland und Südostasien*. Technischer Bericht. Aarau: Fachhochschule Nordwestschweiz Pädagogische Hochschule.
- Judd, Charles H. 1908. "The relation of special training to general intelligence". *Educational review* 36 (28-42).
- Killen, Roy. 2000. *Outcomes-based education: Principles and possibilities*. Technischer Bericht. Faculty of Education, University of Newcastle, Australia.
- Klieme, Eckhard. 2004. "Was sind Kompetenzen und wie lassen sie sich messen?" *Pädagogik* 6:10–13.
- Konsortium HarmoS Naturwissenschaften+. 2010. *Wissenschaftlicher Kurzbericht und Kompetenzmodell*. Technischer Bericht.
- Kultusministerkonferenz. 2004. *Bildungsstandards der Kultusministerkonferenz*.
- LaBerge, David, und S.Jay Samuels. 1974. "Toward a theory of automatic information processing in reading". *Cognitive Psychology* 6, Nr. 2 (April): 293–323.
- Lave, Jean. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*.
- Lersch, Rainer. 2007. "Kompetenzfördernd unterrichten. 22 Schritte von der Theorie zur Praxis". *Pädagogik* 59 (12): 36–43.
- Lobato, Joanne, und Daniel Siebert. 2002. "Quantitative reasoning in a reconceived view of transfer". *The Journal of Mathematical Behavior* 21, Nr. 1 (Januar): 87–116.
- Marini, Anthony, Anne McKeough und Judy Lupart. 1995. *Teaching for Transfer: Fostering Generalization in Learning*. Mallory International.
- Martin, Michael O., und Ina V.S. Mullis. 2003. "Overview of TIMSS 2003". *TIMSS*:2–21.
-

- McGee, Clive. 1996. "The Development of a New National Curriculum in New Zealand". *The Educational Forum* 60, Nr. 1 (März): 56–63.
- Metzger, Susanne, Pitt Hild, Christoph Gut und Josiane Tardent. 2013. "Projekt ExKo-Nawi: Aufgaben und erste Ergebnisse der hands-on Assessments". In *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht*, herausgegeben von Sascha Bernholt, 174–176.
- Michael, Ann L., Thomas Klee, John D. Brandsford und Steven F. Warren. 1993. "The transition from theory to therapy: Test of two instructional methods". *Applied Cognitive ...* 7:139–153.
- Mietzel, Gerd. 2007. *Pädagogische Psychologie des Lernens und Lehrens*. 567. Hogrefe Verlag.
- Millar, Robin, und Jonathan Osborne. 1999. "Beyond 2000: Science education for the future". *Science And Technology*.
- Munier, Valérie, Hélène Merle und Danie Brehelin. 2013. "Teaching Scientific Measurement and Uncertainty in Elementary School". *International Journal of Science Education* 35 (16): 2752–2783.
- Oelkers, Jürgen, Kurt Reusser, Esther Berner, Uli Halbheer und Stefanie Stolz. 2008. *Qualität entwickeln- Standards sichern- mit Differenz umgehen*. Technischer Bericht. Bonn: Pädagogisches Institut der Universität Zürich.
- Pea, Roy D. 2013. "Putting Knowledge to Use". In *Technology in Education: Looking Toward 2020*, herausgegeben von Nickerson Raymond S. und Philop P. Zodiates, 169–212. Routledge.
- Perkins, D.N., und Gavriel Salomon. 1989. "Are cognitive skills context-bound?" *Educational researcher* 18 (1): 16–25.
- PISA-Konsortium Deutschland. 2004. "PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland–Ergebnisse des zweiten internationalen Vergleichs".
- Porter, a. 1989. "A Curriculum out of Balance: The Case of Elementary School Mathematics". *Educational Researcher* 18, Nr. 5 (Juni): 9–15.
- Renkl, Alexander, Hans Gruber, Heinz Mandl und Ludwig Hinkhofer. 1994. "Hilft Wissen bei der Identifikation und Kontrolle eines komplexen ökonomischen Systems?" *Unterrichtswissenschaft* 22:195–202.
- Reusser, Kurt. 2005. "Problemorientiertes Lernen - Tiefenstruktur, Gestaltungsformen, Wirkung". *Beitraege zur Lehrerbildung* 23 (2): 159–182.
- Rindermann, Heiner. 2006. "Was messen internationale Schulleistungsstudien?" *Psychologische Rundschau* 57, Nr. 2 (April): 69–86.
- Rychen, Dominique Simone, und Laura Hersch Salganik. 2001. *Defining and selecting key competencies*. Herausgegeben von Dominique Simone Rychen und Laura Hersch Salganik. 251. Hogrefe & Huber.
-

- Schoenfeld, Alan H. 1988. "When good teaching leads to bad results: The disasters of 'well-taught' mathematics courses". *Educational psychologist* 23 (2): 145–166.
- Schöne, Claudia, Oliver Dickhäuser, Birgit Spinath und Joachim Stiensmeier-Pelster. 2002. *SESSKO Skalen zur Erfassung des schulischen Selbstkonzeptes*. Göttingen: Hoegrefe Verlag.
- Shakeshaft, Nicholas G., Maciej Trzaskowski, Andrew McMillan, Kaili Rimfeld, Eva Krapohl, Claire M a Haworth, Philip S Dale und Robert Plomin. 2013. "Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16." *PloS one* 8, Nr. 12 (Januar): e80341.
- Shuell, Thomas J. 1996. "The role of educational psychology in the preparation of teachers". *Educational Psychologist* 31 (1): 37–41.
- Sichau, David. 2015. "Entwicklung eines ExKoNawi hands-on Test zur skalenbasierten Messung". Forschungsarbeit, PH Zürich.
- Weinert, Franz E. 2001. "Concept of competence: A conceptual clarification." In *Defining and selecting key competencies*, herausgegeben von D. S. Rychen und L. H. Salganik. Seattle: Hogrefe & Huber.
- Whitehead, Alfred North. 1929. *The Aims of Education and Other Essays*, 13–26. New York: Free Press.
- Wiggins, Grant. 1993. "Assessment: Authenticity, context, and validity." *Phi Delta Kappan* 75 (3): 200–208.
- Williams, Susan M. 1992. "Putting case-based instruction into context: Examples from legal and medical education". *The Journal of the Learning Sciences* 2 (4): 367–427.
- Woodworth, Robert S., und Edward L. Thorndike. 1901. "The influence of improvement in one mental function upon the efficiency of other functions. (I)." *Psychological Review* 8:247–261.
-

# Anhang