

TIPP – Teachers in Practice and Process

Schlussbericht über die empirische Evaluation der Bielefelder Referendarausbildung

Prof. Dr. Rainer Dollase, Dipl.Psych. und Dipl.Päd. Odette Selders

Universität Bielefeld, Abt. Psychologie und IKG

Bielefeld, August 2009

Zusammenfassung

Fragestellung: Welche längsschnittlichen Effekte zeigt die Referendarausbildung (insbesondere die Trainingsgruppe) der Bielefelder Gruppe des TIPP-Projektes? Wie unterscheidet sich die Bielefelder Gruppe insgesamt von einer Kontrollgruppe?

Treatment: Die Versuchsgruppe und die darin enthaltene Trainingsgruppe der Bielefelder Referendariatsausbildung legt einen Schwerpunkt auf die Bereiche „Kommunikation“, „Konfliktregelung“ und „Klassenführung“ und bietet ihren Referendaren darin besondere Kurse an. Das unterscheidet sie von der Kontrollgruppe.

Design und Stichprobe: Es wurde ein Kontrollgruppendesign verwendet - zu der Bielefelder Gruppe (N=56) wurde eine Kontrollgruppe (N=55) in der Region gefunden, die ein Ausbildungsprogramm wie bisher üblich absolvierte. In der Versuchsgruppe splittete sich eine Trainingsgruppe ab, die besondere Kurse absolvierte und die insgesamt vier mal untersucht wurde - die übrige Versuchsgruppe insgesamt drei mal. Die Kontrollgruppe wurde zu Anfang und zu Ende der Projektlaufzeit untersucht.

Methode: Es wurde ein Fragebogen eingesetzt, der vier Blöcke von Fragen umfasste. Im ersten Block wurde die Güte der Ausbildung, der Stand des Wissens und des jetzigen Verhaltens in Kommunikation, Konfliktregelung und Klassenführung eingeschätzt. Im zweiten Block wurden Noten für die Wichtigkeit der Ausbildung in Kommunikation, Konfliktregelung und Klassenführung erfragt. Im dritten Block wurde die Wichtigkeit der drei Bereiche - ergänzt um Fachwissen und Unterrichtsablauf - in einem systematischen Paarvergleich erfasst. Im vierten Block wurden Erwartungen künftiger Schwierigkeiten erfragt.

Ergebnisse:

Befund 1: Die Kontrollgruppe bewertet am Ende des Versuchs die Ausbildung, das Wissen und das jetzige Verhalten subjektiv geringfügig besser.

Befund 2: Die Versuchsgruppe bewertet in zentralen Bereichen die kommenden Herausforderungen der Praxis (Disziplin, Konfliktlösung und Stoff verständlich machen) als weniger schwierig - obwohl die Ausgangslage in beiden Gruppen identisch war.

Befund 3: Mit zunehmender Praxis und Ausbildung wird die Wahrnehmung der Wichtigkeit der drei Bereiche Kommunikation, Konfliktregelung und Klassenführung schwächer. Die Unterschiede in der Beurteilungsskala sind nicht sehr dramatisch. Es handelt sich um Unterschiede, die zwischen 0,2 und 0,6 Skalenstufen schwanken.

Befund 4: Mit Hilfe einer besonders zuverlässigen Skalierungsmethode (Paarvergleich) kann nicht nur der o.g. Befund bestätigt werden, dass die subjektiv empfundene Wichtigkeit der ausgewählten Ausbildungsbereiche während des Referendariats sinkt, sondern auch dass die Wichtigkeitsbeurteilung aller näher zusammenrückt. Die eher psychologischen Bereiche Konfliktregelung, Kommunikation und Klassenführung rücken näher an die Beurteilung der fachdidaktischen Bereiche Unterrichtsablauf und Fachwissen heran. Besonders sinkt die Wichtigkeitsbeurteilung der Klassenführung in Versuchs- und Kontrollgruppe.

Befund 5: Bezogen auf Ausbildung, Wissen und Verhalten zeigen die Ergebnisse der Trainingsgruppe innerhalb der Versuchsgruppe über die vier Messzeitpunkte hinweg eine Verbesserung der subjektiven Einschätzung in allen drei Bereichen (Kommunikation, Konfliktregelung und Klassenführung). Der Verlauf ist nicht in allen Fällen kontinuierlich.

Befund 6: Die Trainingsgruppe innerhalb der Versuchsgruppe ist in den meisten Variablen besser als die versuchsgruppeninterne Kontrollgruppe. Das muss bei der Bewertung der Ergebnisunterschiede zwischen Kontrollgruppe und Versuchsgruppe berücksichtigt werden.

Befund 7: Referendare in der Trainingsgruppe bewerten die Schwierigkeiten, Schüler ruhig und diszipliniert zu unterrichten, ebenso wie allen Schülern gerecht werden und den Umgang mit Heterogenität unverändert schwierig. Die Ausbildung hat auf die Erwartungsänderung keinen Einfluss.

Befund 8: Auch in der Trainingsgruppe sinkt die Wichtigkeitseinschätzung von Kommunikation, Konfliktregelung und Klassenführung - bleibt aber noch im „wichtigen“ Bereich. Die Klassenführung sinkt im Paarvergleich am stärksten.

Befund 9: In Versuchsgruppe und Kontrollgruppe hat das jetzige Wissen auf das jetzige Verhalten den stärksten Einfluss - stärker als die Einschätzung der jetzigen Ausbildung. Keinen Einfluss hat die Wichtigkeitseinschätzung (multiple Regression).

Befund 10: Sowohl in der Kontroll- wie auch Versuchsgruppe gibt es einen signifikanten Zusammenhang zwischen den erwarteten Schwierigkeiten „Konflikte schnell und effizient regeln zu können“ und der Selbsteinschätzung des jetzigen „Konfliktregelungsverhaltens“. Das jetzige Klassenführungsverhalten steht in keiner systematischen Beziehungen mit den erwarteten Schwierigkeiten. In der Versuchsgruppe gilt dies auch für die Kommunikation - allerdings nicht in der Kontrollgruppe.

Befund 11: Auch das Anfangswissen, das aus der universitären Ausbildung stammt, hat einen Einfluss auf die Einschätzung des „jetzigen Verhaltens“ in Kommunikation, Konfliktregelung und Klassenführung am Ende des Projektzeitraums.

1. Fragestellung

Die Notwendigkeit einer empirischen Evaluation pädagogischer Programme und Projekte ist heute unbestritten. Zu oft hat sich in der Vergangenheit gezeigt, dass gut gemeinte Ansätze ihre behaupteten Wirkungen nicht erreichen. Daraus hat sich eine gewisse Skepsis gegenüber zwar gut begründeten und gemeinten Programmen entwickelt, die zu der zunehmend häufigeren empirischen, meist auch quantitativen, Evaluation der Effekte geführt hat.

Die Evaluation von Programmen und Projekten in der Lehrerbildung hat einen besonders hohen Stellenwert, da in der internationalen empirischen Unterrichtsforschung den Fähigkeiten der Lehrpersonen ein besonders hoher Stellenwert bescheinigt wird. Zigtausende Untersuchungen beschäftigen sich damit (Wang u.a., 1993).

Die Effektmessung als solche ist ein umfängliches wissenschaftliches Gebiet und keineswegs unproblematisch. Selten sind die Verfahren, die eingesetzt werden, so sensibel, dass sie auch kleinste Veränderungen nachzeichnen können, und andererseits ist man immer wieder im so genannten „Bandbreite-Fidelitäts-Dilemma“ gefangen, d.h. eine ausführliche und umfangreiche Befragung wäre eigentlich günstig, würde sich aber spezialisieren müssen und die Breite möglicher Effekte nicht erfassen können. Solche und ähnliche Probleme sind seit Jahrzehnten bekannt und stellen sich auch in dieser Untersuchung (Lösel u.a., 1987).

Von besonderer Bedeutung ist bei einer Evaluation einmal die Längsschnittlichkeit der Untersuchung und zum anderen die Existenz einer Kontrollgruppe. Wirkungsevaluation kann nur über Längsschnitte erfolgen, da man nachweisen muss, dass sich die Parameter im Laufe des Trainings oder des Programms gebessert haben. Eine Kontrollgruppe ist notwendig, damit der Vorteil des Programms oder Trainings gegenüber den herkömmlichen bzw. gegenüber anderen Ansätzen erwiesen werden kann.

Im TIPP-Projekt heißt das, dass folgende Fragestellungen beantwortet werden müssen:

1. Welche längsschnittlichen Effekte zeigt das Trainingsprogramm der Bielefelder Gruppe des TIPP-Projektes?
2. Wie unterscheidet sich die Bielefelder Gruppe insgesamt von einer Kontrollgruppe?
- 3.

Für den Schlussbericht der Evaluation ist insbesondere interessant, ob die Ausgangslagen zwischen Trainingsgruppe und Kontrollgruppe vergleichbar waren, denn wenn Ausgangsunterschiede existieren, müssten diese durch eine so genannte Kovarianzanalyse (Ausgangslage als Kovariat) kontrolliert werden.

Das Bandbreite-Fidelitäts-Dilemma wird durch einen kurzen, d.h. nur zweiseitigen Fragebogen gelöst, d.h. die Fragen und Items sind auf globale Einschätzungen der 3 inhaltlich wichtigen Bereiche „Kommunikation“, „Konfliktbearbeitung“ und „Klassenführung“ bezogen.

Genau diese drei Kompetenzbereiche werden in der Bielefelder Trainingsgruppe angeboten. Die Bielefelder Gruppe des TIPP-Projektes hat sich erklärt, dies zu einem Schwerpunkt der Ausbildung zu machen. Die Aufnahme der drei Bereiche im Fragebogen erfüllt die Forderung nach „curricularer Validität“ eines Evaluationsfragebogens, d.h. er soll das erfragen bzw. messen, was auch Anlass und Ziel des Trainings ist.

Treatment: Die Versuchsgruppe und die darin enthaltene Trainingsgruppe der Bielefelder Referendariatsausbildung legt einen Schwerpunkt auf die Bereiche „Kommunikation“, „Konfliktregelung“ und „Klassenführung“ und bietet ihren Referendaren darin besondere Kurse an. Das unterscheidet sie von der Kontrollgruppe.

Die Auswahl der Bereiche Kommunikation, Konfliktbearbeitung und Klassenführung entspricht in gewisser Weise dem aktuellen Stand der empirischen Unterrichtsforschung. Wie die bekannte Metaanalyse von Wang u.a. gezeigt hat, ist das „Klassenführungsverhalten“ (classroom management) der wirksamste Beitrag zum guten Unterricht (kognitive und soziale Erfolgskriterien). Ein angemessenes

Klassenführungsverhalten kann man allerdings nur über eine entsprechende Kommunikation erreichen. „Kommunikation“ ist also ein Bestandteil des Klassenführungsverhalten. Die „Konfliktbearbeitung“ andererseits ist ein Teil der Kommunikation, das heißt, es ist eine Kommunikation in einem speziellen Falle, dem Fall eines Konfliktes zwischen Schülern oder Schülern und Lehrern. Ob man nun Kommunikation als den Oberbegriff für alle drei Bereiche nimmt, oder aber das Klassenführungsverhalten als allgemeinste Kategorie des Lehrerverhaltens vorschlägt, ist nicht ohne Folgen für die Ausfüllung des Fragebogens. Wie noch zu zeigen sein wird, haben viele Referendare die entsprechenden Grenzen und spezifischen Beiträge der einzelnen Begriffe gegen Ende ihrer Referendarzeit offenbar als ähnlich angesehen, beziehungsweise die Begriffe nicht mehr weiter differenziert.

Das ist kein Manko der Ausbildung: das ist der Tatsache einer unscharfen Verwendung der einzelnen Begriffe in der Fachliteratur geschuldet. So sind neue Werke zum Klassenführungsverhalten mittlerweile so unspezifisch, das eine Vermischung der drei Begriffe Kommunikation, Konfliktregelung und Klassenführungsverhalten zwangsläufig daraus folgt (Evertson u.a., 2006; Eichhorn, 2008; Dollase 1995).

2. Stichprobe und Methode

Design und Stichprobe

In Tab. 1 ist neben dem Design erkenntlich, welche Sollstärke für die unterschiedlichen Gruppen im Rahmen des Untersuchungsdesigns vorgesehen waren. Die Versuchsgruppe besteht aus den Bielefelder Referendaren und die Kontrollgruppe wurde in Detmold gefunden. Insgesamt sollten 56 Personen in der Versuchsgruppe und 55 Personen in der Kontrollgruppe sein. Die Versuchsgruppe splittete sich in zwei Gruppen: 22 Personen mit einem gesonderten Konflikt-Bearbeitungs-Training und 34 Personen, die dieses Training nicht erhalten haben. Deshalb kann man in dieser Studie auch von einer „versuchsgruppeninternen“ Kontrollgruppe sprechen.

Tabelle 1: Sollstärken der Stichproben und Design der Untersuchung. Sollstärke der Gesamtstichprobe 111.

Termine	Versuchsgruppe		Kontrollgruppe
	Training	kein Training	
1	22	34	55
2	22	34	-
3	22	-	-
4	22	34	55

In Tab. 2 ist die Ist-Stärke der Stichproben erkenntlich. Die Trainingsgruppe konnte in den Messungen drei und vier auf die Sollstärke aufgefüllt werden, die versuchsgruppeninterne Kontrollgruppe wurde im Verlaufe der Untersuchung geringer. Die Kontrollgruppe erreichte weder bei der ersten noch bei der letzten Messung ihrer Sollstärke.

Tabelle 2: Ist-Stärke der Stichproben, Kriterium: Angabe des Geschlechts.

Gesamtstichprobe t1: 98; t4: 91, Versuchsgruppe t1: 51; t4: 48

Termine	Versuchsgruppe		Kontrollgruppe
	Training	kein Training	
1	18	33	47
2	19	34	-
3	22	-	-
4	22	26	43

Die dritte Erhebung der Versuchsgruppe fand nur für die Trainingsgruppe statt. Deshalb fehlt für diesen Termin eine versuchsgruppeninterne Kontrollgruppe. Die

letzte Erhebung der Kontrollgruppe fand zeitlich in etwa zugleich mit der vierten Erhebung in der Versuchsgruppe statt. Die erste Erhebung fand kurz nach dem Eintritt in das Referendariat statt, die zweite und die dritte Erhebung innerhalb von 4 Monaten nach Beginn des Referendariats.

Die Gründe für die Abweichungen von der Soll Stärke sind vielfältig, dennoch üblich: organisatorische Gründe und personeller Wechsel (Zu- und Abgang) waren wohl am häufigsten.

Design: Es wurde ein Kontrollgruppendesign verwendet - zu der Bielefelder Gruppe wurde eine Kontrollgruppe in der Region gefunden, die ein Ausbildungsprogramm wie bisher üblich absolvierte. In der Versuchsgruppe splittete sich eine Trainingsgruppe ab, die besondere Kurse absolvierte und die insgesamt vier mal untersucht wurde - die Versuchsgruppe insgesamt drei mal. Die Kontrollgruppe wurde zu Anfang und zu Ende der Projektlaufzeit untersucht.

Methoden

Der Fragebogen enthielt neben der Anrede der Referendare und Referendarinnen und der Garantie einer anonymen Beantwortung und Auswertung eine Codierung, die notwendig wurde, weil es sich ja um eine anonyme Längsschnittstudie handelt. Als Code wurde der Geburtsmonat der Mutter, der zweite Buchstabe im Vornamen des Vaters und der letzte Buchstabe des Geburtsortes verlangt. Dadurch war es möglich, in allen Fällen die Fragebögen zuzuordnen.

Im *ersten Fragebogenblock* waren neun Items enthalten, die jeweils für Kommunikation, Konfliktbearbeitung und Klassenführung drei Beurteilungen verlangten und zwar sollte die „Güte der Ausbildung“, das „heutige Wissen“ und das „jetzige Verhalten“ in den drei Bereichen mit einer Schulnote von 1 bis 6 beurteilt werden. Dadurch wird es möglich, subjektive Einschätzungen von Ausbildung, Wissen und Verhalten über die Erhebungszeitpunkte zu erfassen.

Im *zweiten Block* wurden Noten für die Wichtigkeit der drei Bereiche Kommunikation, Konfliktbearbeitung und Klassenführung erfragt (3 Items).

Im *dritten Block* wurde ein vollständiger Paarvergleich (sog. pair comparison technique) verlangt. Und zwar sollte das „Fachwissen in den Fächern“, die „Verbesserung der Kommunikation“, die „Verbesserung der Konfliktbearbeitung“, die „Verbesserung der Klassenführung“, die „Unterrichtsabläufe in den Fächern“ in allen Zweierkombinationen beurteilt werden. Die Frage lautete: „Entscheiden Sie sich jeweils für eine von zwei Alternativen, was ist für die Praxis wichtiger?“ Dann folgten alle möglichen Zweierkombinationen, z.B. Verbesserung meiner Konfliktbearbeitung oder Verbesserung meiner Klassenführung. Mit Hilfe des Paarvergleichsverfahrens kann man besonders zuverlässige Rangreihen der Wichtigkeit erfassen.

Im *vierten Block* wurden die Erwartungen der Referendare bezüglich der größten Herausforderungen in der Praxis erfragt. Die Referendare und Referendarinnen sollten die Höhe der Schwierigkeitserwartung in Schulnoten von 1 = geringe Schwierigkeiten bis 6 = allergrößte Schwierigkeiten angeben. Hier wurden insgesamt 6 (in der zweiten und dritten Messung 7) Themen genannt. Beispiele: Den Schülern/ Schülerinnen den Stoff verständlich vermitteln können. Allen Schülern gerecht zu werden. Der Umgang mit Heterogenität der Schülerschaft.

Methode: Es wurde ein Fragebogen eingesetzt, der vier Blöcke von Fragen umfasste. Im ersten Block wurde die Güte der Ausbildung, der Stand des Wissens und des jetzigen Verhaltens in Kommunikation, Konfliktregelung und Klassenführung eingeschätzt. Im zweiten Block wurden Noten für die Wichtigkeit der Ausbildung in Kommunikation, Konfliktregelung und Klassenführung erfragt. Im dritten Block wurde die Wichtigkeit der drei Bereiche - ergänzt um Fachwissen und Unterrichtsablauf - in einem systematischen Paarvergleich erfasst. Im vierten Block wurden Erwartungen künftiger Schwierigkeiten erfragt.

Zum Schluss wurde noch nach Geschlecht und Alter gefragt, auch nach der fachwissenschaftlichen Orientierung, ob sich der Referendar/die Referendarin eher naturwissenschaftlich, eher gesellschaftswissenschaftlich oder eher sprachwissenschaftlich orientiert.

Der Fragebogen ist im Anhang dieses Berichtes enthalten.

3. Ergebnisse

3.1. Die Prüfung von Anfangsunterschieden zwischen Kontrollgruppe und Trainingsgruppe

Für den Nachweis von Effekten des Trainings ist die Prüfung der Ausgangsunterschiede zwischen Trainings- und Kontrollgruppe wichtig.

Ausbildung, Wissen und Verhalten

In Tabelle 3 sind die entsprechenden Mittelwerte bei Versuchs- und Kontrollgruppe zum ersten und letzten Messzeitpunkt aufgeführt. Es gibt nur zwei signifikante Unterschiede im ersten Messzeitpunkt, und zwar im „Verhalten Kommunikation“, hier ist die Kontrollgruppe besser, und im „Verhalten Klassenführung“, auch hier fühlt sich die Kontrollgruppe etwas besser. Der erste Befund ist auf dem 2%-Niveau, der zweite auf dem 5%-Niveau signifikant.

Tabelle 3: Beurteilung von Ausbildung, Wissen und Verhalten. Mittelwerte zu Anfang und Ende des Untersuchungszeitraums für Versuchs- und Kontrollgruppe

	VG Anfang	VG Ende	KG Anfang	KG Ende
Ausbildung Kommunikation	3,7	3,0	3,3	2,4
Ausbildung Konfliktbearbeitung	4,4	3,3	4,1	2,9
Ausbildung Klassenführung	4,5	3,3	4,4	3,2
Wissen Kommunikation	3,1	2,6	2,9	2,5
Wissen Konfliktbearbeitung	3,6	3,1	3,3	2,8
Wissen Klassenführung	4,0	3,2	3,8	2,8
Verhalten Kommunikation	3,0	2,7	2,6	2,4
Verhalten Konfliktbearbeitung	3,2	3,0	2,9	2,8
Verhalten Klassenführung	3,8	3,0	3,4	2,8

Die obige Tabelle zeigt ja den Vergleich zwischen Versuchs- und Kontrollgruppe (wie erwähnt) am Anfang und auch am Ende des Versuches. Man erkennt unter Berücksichtigung der Endmessung für die Versuchsgruppe: eine Verbesserung sämtlicher Variablen um stellenweise fast eine Notenstufe. Für die Kontrollgruppe ist ebenfalls eine durchgängige Verbesserung sämtlicher Bewertungen von Variablen von Anfang bis Ende festzustellen. Interessant sind Vergleiche zwischen den Endwerten der Versuchsgruppe und den Endwerten der Kontrollgruppe. Hier zeigt sich durchgängig, dass die Kontrollgruppe gegen Ende subjektiv bessere Bewertungen abgibt als die Versuchsgruppe.

Befund: Die Kontrollgruppe bewertet am Ende des Versuchs die Ausbildung, das Wissen und das jetzige Verhalten subjektiv geringfügig besser.

Auf eine Signifikanzprüfung der Unterschiede gegen Ende wurde bewusst verzichtet. Es sind für die Bewertung der Variablen Schulnoten verwendet worden. Es ist anzunehmen, dass bei Lehrern und Fachleitern, aber auch bei den Referendaren einen Notenunterschied von 0,3 Relevanz besitzt. Dieser Unterschied entspricht der Unterteilung der Notenskala und wird auch hier als Bezeichnung für die Relevanz von Unterschieden verwendet.

Erwartungen von Schwierigkeiten in der Praxis

In der folgenden Tabelle werden die erwarteten Schwierigkeiten in der Schulpraxis von den Referendaren der Versuchs- und der Kontrollgruppe in Noten bewertet. Es zeigt sich, trotz der besseren Beurteilung der Fähigkeiten in der Kontrollgruppe, dass die Referendare der Versuchsgruppe einige Fragen als weniger schlimm beantworten. Dies betrifft insbesondere die Aufgabe „Schüler ruhig zu halten“, „Konflikte effektiv zu lösen“ und den „Stoff verständlich“ zu machen. Die Referendare der Kontrollgruppe beurteilen am Ende die Fähigkeit „den richtigen Ton im Umgang mit den Schülern zu treffen“ als weniger schwierig ebenso wie das „notwendige Fachwissen zur Verfügung zu haben“.

Tabelle 4: Erwartete größte Herausforderung in der Praxis – von 1 = geringe bis 6 = größte Schwierigkeiten. Unterschiede zwischen Versuchs- und Kontrollgruppe in der 1. Messung sind sämtlich nicht signifikant

	VG Anfang	VG Ende	KG Anfang	KG Ende
Schüler ruhig halten	3,4	2,9	3,4	3,3
Fachwissen haben	3,0	2,8	2,7	2,4
richtigen Ton treffen	2,6	2,3	2,4	1,9
Konflikte effektiv lösen	3,2	2,9	3,3	3,2
Stoff verständlich machen	3,0	2,6	2,8	2,9
Allen gerecht werden	3,7	3,5	3,9	3,6

Für die Ausgangsmessung wurde gefunden, dass alle Items nicht signifikant sind, d.h. es gibt keine Unterschiede zwischen Versuchs- und Kontrollgruppe. Eine entsprechende Diskriminanzanalyse ermittelt demzufolge auch nur 57% korrekt klassifizierte Fälle, d.h. 7% mehr als Zufall.

Die absoluten Notenwerte sind auch interessant: „den richtigen Ton treffen“ wird als nicht so schwierig wie z.B. die Forderungen „Schüler ruhig halten“ oder „Allen gerecht werden“ beurteilt.

Befund: Die Versuchsgruppe bewertet in zentralen Bereichen die kommenden Herausforderungen der Praxis (Disziplin, Konfliktlösung und Stoff verständlich machen) als weniger schwierig - obwohl die Ausgangslage in beiden Gruppen identisch war.

Um eine multivariate Prüfung der Anfangsunterschiede durchzuführen wurde eine Diskriminanzanalyse gerechnet. Hierbei wurden alle Ausbildungs-, Wissens- und

Verhaltensfragen des Fragebogenblocks als abhängige Variablen in die Diskriminanzanalyse zur Unterscheidung von Gruppe 1 (Trainings-/Versuchsgruppe) und Gruppe 2 (Kontrollgruppe) eingegeben. Das Klassifikationsresultat ist nicht sehr deutlich. 61% der Fälle, also nur 11% mehr als Zufall, werden korrekt klassifiziert.

Beurteilung der Wichtigkeit von Kommunikation, Konfliktregelung und Klassenführung

Die Beurteilung der Wichtigkeit einer Verbesserung von Kommunikation, Konfliktregelung und Klassenführung kann man ebenfalls am Ende und Anfang eines Experimentes durch eine einfache Ratingskala erheben. In der Veränderung der Wichtigkeit- so die Annahme - spiegelt sich die Veränderung der Beurteilung von Kompetenzen im Lehrerberuf wieder. Man erkennt bei einer einfachen Skala der Wichtigkeitsnoten, dass die drei Bereiche anfänglich mit besseren Noten beurteilt worden sind als gegen Ende.

Tabelle 5: Wichtigkeitsveränderungen in Versuchs- und Kontrollgruppe. Dargestellt sind die Werte je für Anfangs- und Endmessung- gemessen mit einer Rating Skala.

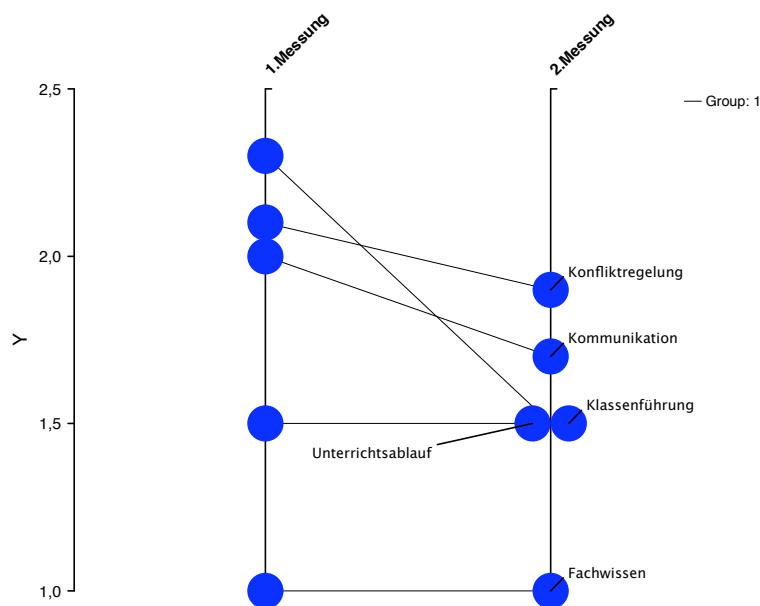
Verbesserung der..	Versuchsgruppe		Kontrollgruppe	
	Anfang	Ende	Anfang	Ende
Kommunikation	1,5	1,8	1,8	2,0
Konfliktregelung	1,4	1,6	1,4	1,9
Klassenführung	1,4	1,6	1,5	2,1

Befund: Mit zunehmender Praxis und Ausbildung wird die Wahrnehmung der Wichtigkeit der drei Bereiche Kommunikation, Konfliktregelung und Klassenführung schwächer. Die Unterschiede in der Beurteilungsskala sind nicht sehr dramatisch. Es handelt sich um Unterschiede, die zwischen 0,2 und 0,6 Skalenstufen schwanken.

Wie bekannt, sind Beurteilungsskalen verschiedenartigen Verzerrungen ausgesetzt (z.B. Milde- und Strengfehler, Halo Effekt). Um diese auszuschalten, wendet man

den „systematischen Paarvergleich“ an, bei dem die Versuchspersonen alle möglichen Zweier-Paarungen der Variablen beurteilen sollen. Sie sollen jeweils angeben, welcher von den beiden Alternativen wichtiger ist als die andere. In der folgenden Abbildung ist das Ergebnis eines systematischen Paarvergleichs zwischen den Themenbereichen „Kommunikation“, „Konfliktregelung“, „Klassenführung“, „Unterrichtsablauf“ und „Fachwissen“ dargestellt. In der Abbildung 1 ist das Ergebnis für die Versuchsgruppe, in der Abbildung 2 das Ergebnis für die Kontrollgruppe dargestellt.

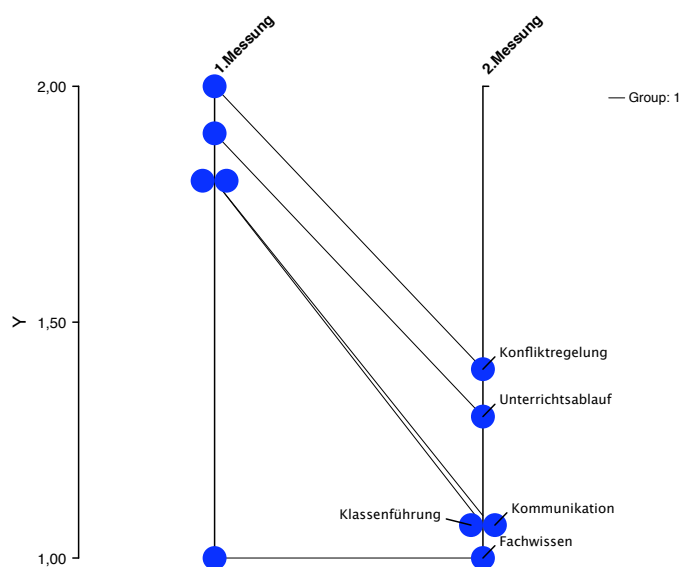
Abbildung 1: Ergebnis des systematischen Paarvergleichs nach Wichtigkeit der Ausbildungsfelder für die **Versuchsgruppe** - Skalenwert zeigt das Mehrfache des Wertes für Fachwissen



Man kann nun erkennen, dass die Wichtigkeitsbeurteilungen von der ersten zur zweiten Messung näher aneinander rücken und insgesamt (siehe oben) sinken. Das bedeutet: am Ende des Referendariats ist die Rangreihe der Prioritäten nicht mehr so weit auseinander gezogenen wie am Anfang. In der Versuchsgruppe erkennt man ebenfalls eine deutliche Veränderung der Rangreihe. Führte die Klassenführung noch deutlich bei der ersten Messung, so liegt sie in der zweiten Messung

zusammen mit dem Unterrichtsablauf auf dem zweitletzten Platz. Aufgestiegen ist keine Variable. Dieser Aussage kann man auch für die Kontrollgruppe treffen. Auch hier ist keine Variable aufgestiegen, sondern alle sinken in ihrem Skalenwert. Auch in der Kontrollgruppe ist Klassenführung, Kommunikation und Fachwissen nicht mehr so wichtig wie Konfliktregelung und Unterrichtsablauf.

Abbildung 2: Ergebnis des systematischen Paarvergleichs nach Wichtigkeit der Ausbildungsfelder für die **Kontrollgruppe** - Skalenwert zeigt das Mehrfache des Wertes für Fachwissen



Befund: Mit Hilfe einer besonders zuverlässigen Skalierungsmethode (Paarvergleich) kann nicht nur der o.g. Befund bestätigt werden, dass die subjektiv empfundene Wichtigkeit der ausgewählten Ausbildungsbereiche während des Referendariats sinkt, sondern auch dass die Wichtigkeitsbeurteilung aller näher zusammenrückt. Die eher psychologischen Bereiche Konflikt, Kommunikation und Klassenführung rücken näher an die Beurteilung der fachdidaktischen Bereiche Unterrichtsablauf und Fachwissen heran. Besonders sinkt die Wichtigkeitsbeurteilung der Klassenführung in Versuchs- und Kontrollgruppe.

Zum Verständnis dieser Ergebnisse ist eine Kenntnis der Skalierungsmethode Paarvergleich notwendig. Durch die Paarvergleiche können in einer Matrix Gewinnerprozentsätze (Variable A wird gegenüber Variable B in z.B. 54% der Fälle vorgezogen) für jede Variable festgestellt werden. Diese werden addiert, so dass es eine eindeutige Rangreihe nach Prozenten des Gewinns gegen konkurrierende Variablen gibt. Diese Summenprozentsätze kann man nun skalieren. In unserem Falle wird der Prozentsatz für das Fachwissen, das auf dem letzten Platz gelandet ist (also gegen alle anderen Variablen verloren hat) gleich eins gesetzt. Die Skalenwerte der anderen Variablen ergeben sich nun, wie in einer Rationalskala, als Mehrfaches des Wertes für Fachwissen. Das Zusammenrücken der einzelnen Variablen in der Wichtigkeitsbeurteilung zeigt indirekt auch, dass die Unwichtigkeit des Fachwissens am Ende nicht mehr so dramatisch wie am Anfang wahrgenommen wird.

3.2 Längsschnittliche Unterschiede in der Trainingsgruppe

Die Erwartungen des TIPP-Projektes für die Trainingsgruppe (innerhalb der Versuchsgruppe) in Bielefeld sind, dass sich über die Messzeitpunkte Verbesserungen in der Einschätzung von Ausbildung, Wissen und Verhalten in den drei zentralen Inhaltsbereichen ergeben. Hiermit kovariierend sollte die Beurteilung der Herausforderung unterschiedlicher Praxisprobleme sinken und die Wichtigkeitsbeurteilung einer Verbesserung sich ebenfalls verändern (wie ist nicht weiter bestimmt worden).

In Tabelle 6 sind die Ergebnisse über die Zeitpunkte der Trainingsgruppe in Ausbildung, Wissen und Verhalten dargestellt. T1 wird gegen t3 geprüft und man erkennt, dass die Ausbildung in Kommunikation, Konfliktbearbeitung und Klassenführung deutlich besser geworden ist (signifikant). Ebenfalls gilt dies für das „jetzige Wissen“ in Kommunikation und Konfliktbearbeitung, auch das wird bei t3 signifikant besser beurteilt, nicht aber das Wissen im Bereich Klassenführung, hier ergibt sich ein nicht signifikantes Ergebnis. Gegenüber t4 erfolgte keine Signifikanzprüfung mehr (vgl. Tabelle 7).

Tabelle 6: Ausbildung, Wissen und Verhalten. Entwicklung in der Trainingsgruppe über vier Zeitpunkte.

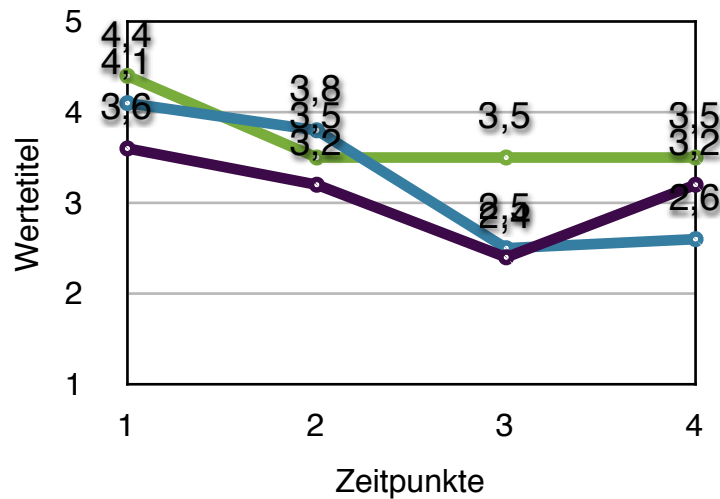
	t1	t2	t3	t4
Ausbildung Kommunikation	3,6	3,2	2,4	3,2
Ausbildung Konfliktbearbeitung	4,1	3,8	2,5	2,6
Ausbildung Klassenführung	4,4	3,5	3,5	3,5
Wissen Kommunikation	2,9	2,7	2,2	2,4
Wissen Konfliktbearbeitung	3,1	3,1	2,6	2,4
Wissen Klassenführung	3,6	3,2	3,4	2,8
Verhalten Kommunikation	2,8	2,6	2,6	2,4
Verhalten Konfliktbearbeitung	2,9	2,9	2,8	2,5
Verhalten Klassenführung	3,5	3,4	3,1	2,7

Weil angenommen werden kann, dass die Tabelle 6 von großer Bedeutung für die Bewertung des Projektes ist, sind die entsprechenden Werte in den folgenden drei Abbildungen grafisch dargestellt worden, um einen visuellen Eindruck von den Veränderungen deutlich zu machen.

In der ersten Abbildung (3) ist der Verlauf der Bewertung der Ausbildung erkenntlich. Punktuell sieht man, dass die Noten sich von der ersten zur letzten Messung verbessern, auch wenn die Ausbildung in Kommunikation und ein wenig auch die Ausbildung in Konfliktbearbeitung bei der dritten Messung einen Optimum hatten, das gegen Ende wieder etwas verschlechtert wird. Die Ausbildung in Klassenführung stagniert seit der zweiten Messung auf dem Wert 3,5.

Abbildung 3: Trainingsgruppe innerhalb der Versuchsgruppe. Noten für die Ausbildung in den drei verschiedenen Themen. (Niedrige Werte sind bessere Werte!)

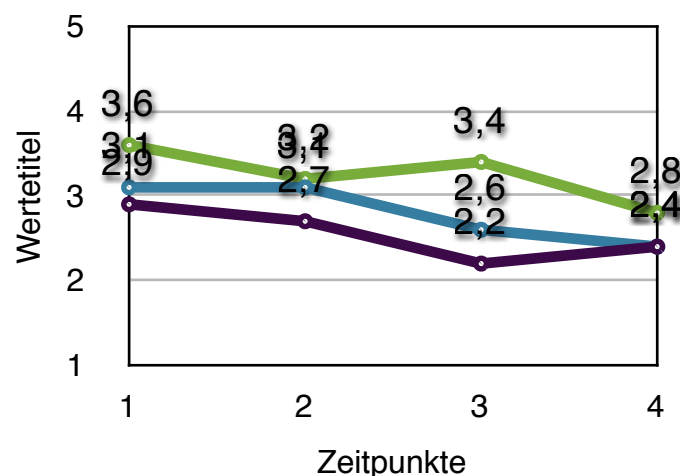
- Ausbildung Kommunikation
- Ausbildung Konfliktbearbeitung
- Ausbildung Klassenführung



In der zweiten Abbildung (4) wird der Verlauf der Wissens einschätzung dargestellt. Man erkennt, dass hier leichte Verbesserungen in allen drei Variablen zu konstatieren sind.

Abbildung 4: Trainingsgruppe innerhalb der Versuchsgruppe. Noten für das Wissen in den drei verschiedenen Themen (Niedrige Werte sind bessere Werte!)

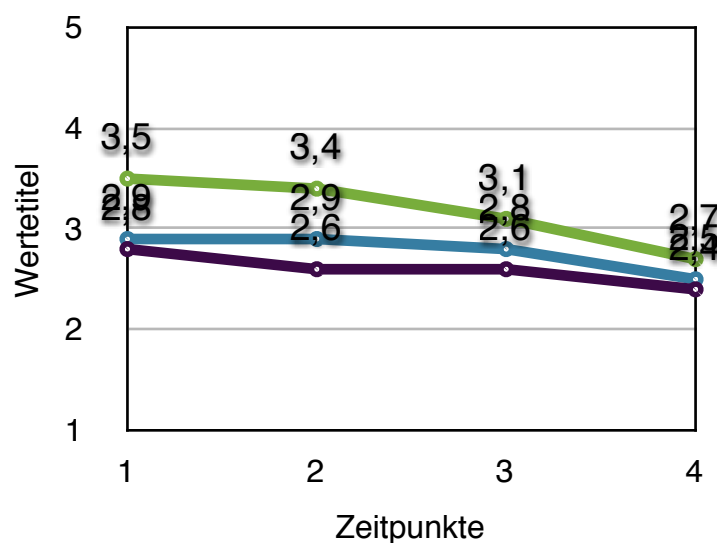
- Wissen Kommunikation
- Wissen Konflikte bearbeiten
- Wissen Klassenführung



In der dritten Abbildung (5) erkennt man, dass sich das Verhalten in Kommunikation, Konfliktbearbeitung und Klassenführung im Laufe des Referendariats ein wenig geändert hat.

Abbildung 5: Trainingsgruppe innerhalb der Versuchsgruppe. Noten für das jetzige Verhalten in den drei verschiedenen Themen (Niedrige Werte sind bessere Werte!)

- Verhalten Kommunikation
- Verhalten Konfliktbearbeitung
- Verhalten Klassenführung



Man kann also für alle Bereiche eine Verbesserung der Notenbewertung von Ausbildung, Wissen und Verhalten durch die Referendare feststellen. Es ist hier noch einmal darauf hinzuweisen, dass es sich um subjektive Daten handelt, die man mithilfe der hier gewählten Methode nicht objektivieren kann.

Befund: Bezogen auf Ausbildung, Wissen und Verhalten zeigen die Ergebnisse der Trainingsgruppe innerhalb der Versuchsgruppe über die vier Messzeitpunkte hinweg eine Verbesserung der subjektiven Einschätzung in allen drei Bereichen (Kommunikation, Konfliktregelung und Klassenführung). Der Verlauf ist nicht in allen Fällen kontinuierlich.

Eine besondere Prüfung ist innerhalb der Versuchsgruppe möglich. In der Tabelle 7 wird die Trainingsgruppe mit der Gruppe ohne Training verglichen. Und zwar wurde die letzte Messung bei t4 als Grundlage für die Auswertung verwendet. Man erkennt, dass in vielen Variablen die Trainingsgruppe besser ist: insgesamt in acht Variablen. Die versuchsgruppeninterne Kontrollgruppe ist lediglich in vier Variablen besser. Die restlichen Variablen zeigen, gemessen am Unterschiedswert von 0,3 Notenstufen, ein ähnliches Ausprägungsbild.

Qualitativ hat offenbar der Kurs in Konfliktregelung (der Trainingsgruppe) die Kommunikation ebenfalls verbessert, was insgesamt dazu geführt hat, dass die Trainingsgruppe besser ist als die versuchsgruppeninterne Kontrollgruppe.

Tabelle 7: Trainingsgruppe und Gruppe ohne Training in der Versuchsgruppe (versuchsgruppeninterne Kontrollgruppe). Vergleich bei der **letzten** Messung. Ab 0,3 Notenstufe Unterschied wird das Urteil „besser“ vergeben (Niedrige Werte sind bessere Werte)

Variable	Training t4	kein Training t4	Ergebnis
Ausbildung Kommunikation	3,2	3,4	T und K ähnlich
Ausbildung Konfliktregelung	2,6	3,8	T besser
Ausbildung Klassenführung	3,5	3,1	K besser
Wissen Kommunikation	2,4	2,9	T besser
Wissen Konfliktregelung	2,4	3,5	T besser
Wissen Klassenführung	2,8	3,0	T und K ähnlich
Verhalten Kommunikation	2,4	2,7	T besser
Verhalten Konfliktregelung	2,5	3,3	T besser
Verhalten Klassenführung	2,7	3,0	T besser
Wichtigkeit Kommunikation	2,0	1,9	T und K ähnlich
Wichtigkeit Konfliktregelung	2,0	1,5	K besser
Wichtigkeit Klassenführung	1,8	1,5	K besser
Schwierigkeit Disziplin	3,3	3,1	T und K ähnlich
Schwierigkeit Fachwissen	2,4	3,0	T besser
Schwierigkeit richtiger Ton	2,0	2,3	T besser
Schwierigkeit Konfliktregelung	2,9	3,1	T und K ähnlich
Schwierigkeit Stoff verständlich machen	2,7	2,8	T und K ähnlich
Schwierigkeit allen gerecht werden	4,0	3,5	K besser

Befund: Die Trainingsgruppe innerhalb der Versuchsgruppe ist in den meisten Variablen besser als die versuchsgruppeninterne Kontrollgruppe. Das muss bei der Bewertung der Ergebnisunterschiede zwischen Kontrollgruppe und Versuchsgruppe berücksichtigt werden.

Schwierigkeitserwartungen in der Trainingsgruppe

In der nächsten Tabelle (8) sind die Ergebnisse für die Schwierigkeitserwartungen zu vier Erhebungszeitpunkten für die Trainingsgruppe dargestellt. Und zwar für eine Reihe von praxisnahen Variablen. Hierbei ist erkenntlich, dass einige dieser Variablen sich nicht verbessern: „die Schüler ruhig zu halten“, die Aufgabe „allen gerecht zu werden“, „angemessen mit Heterogenität“ umzugehen - das sind Bereiche, die während des Referendariats kaum entschärft werden konnten. Interessant ist, dass „Konflikte lösen“ sich nur von 3,1 auf 2,9 verbessert, also trotz eines als positiv bewerteten Trainings seinen „Schrecken“ nicht verliert. Ebenfalls theoretisch interessant ist, dass sich die Schwierigkeit „Schüler ruhig und diszipliniert zu halten“ - einer Domäne des Klassenführungsverhaltens - ebenfalls kaum entproblematisiert.

Tabelle 8: Erwartung praktischer Schwierigkeiten in der Trainingsgruppe.

	t1	t2	t3	t4
Schüler ruhig halten	3,5	3,2	2,8	3,3
Fachwissen haben	2,8	2,7	2,5	2,4
richtigen Ton treffen	2,4	2,2	2,3	2,0
Konflikte effektiv lösen	3,1	3,3	2,8	2,9
Stoff verständlich machen	2,9	2,6	2,4	2,7
Allen gerecht werden	3,7	3,6	3,4	4,0
Heterogenität	-	3,6	3,4	3,5

Befund: Referendare in der Trainingsgruppe bewerten die Schwierigkeiten, Schüler ruhig und diszipliniert zu unterrichten, ebenso wie allen Schülern gerecht werden und den Umgang mit Heterogenität unverändert schwierig. Die Ausbildung hat auf die Erwartungsänderung keinen Einfluss.

Wichtigkeitsbeurteilung in der Trainingsgruppe

In der folgenden Tabelle (9) ist die Veränderung der Wichtigkeitsbeurteilung in der Trainingsgruppe dargestellt. Man sieht den auch weiter oben festgestellten Trend zu einer sinkenden Wichtigkeitsbeurteilung von Kommunikation, Konfliktbearbeitung und Klassenführung. Die Wichtigkeitsbeurteilung der Verbesserung von Kommunikation, Konfliktbearbeitung und Klassenführung sinkt über die vier Zeitpunkte in der Trainingsgruppe– in den Bereichen zwischen 0,5 und 0,6 Skalenstufen.

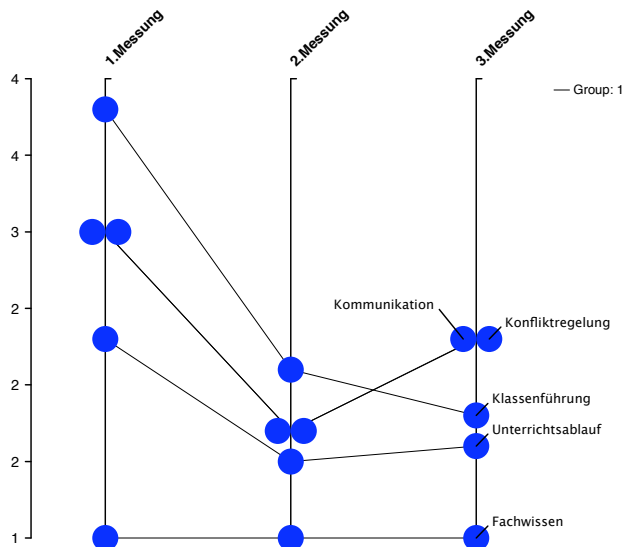
Möglicherweise ergibt sich durch die intensive Auseinandersetzung mit der Praxis ein geschärftes Bewusstsein für die multifaktorielle Determiniertheit des guten Unterrichts, in dessen Gefolge die für die Ausbildung ausgewählten Schwerpunktbereiche Kommunikation, Konfliktbearbeitung und Klassenführung weniger wichtig erscheinen. Möglich ist aber auch, dass man in der Auseinandersetzung mit der Praxis erkennt, dass die zu erzielenden Verbesserungen dadurch zu gering sind, um ein wirkliches Erfolgserlebnis im Unterricht zu haben. Oder: die Realisierung der Ausbildung in den Bereichen entsprach nicht den (überhöhten) Wichtigkeitserwartungen.

Tabelle 9: Wichtigkeitsbeurteilung in der Trainingsgruppe über die vier Zeitpunkte

	t1	t2	t3	t4
Wichtigkeit Verbesserung Kommunikation	1,4	1,6	1,7	2,0
Wichtigkeit Verbesserung Konfliktbearbeitung	1,3	1,6	1,7	2,0
Wichtigkeit Verbesserung Klassenführung	1,3	1,6	1,8	1,8

Für die Trainingsgruppe wurde ebenfalls ein systematischer Paarvergleich der Wichtigkeitsbeurteilung durchgeführt. Auch hierbei zeigt sich, dass die Rangreihe der Wichtigkeit von Kommunikation, Konfliktregelung, Klassenführung, Unterrichtsablauf und Fachwissen sich von der ersten zur zweiten Messung entspreizt und von der zweiten zur dritten Messung wieder ein wenig auseinandergeht. Die Klassenführung verliert ihren ersten Platz von der zweiten zur dritten Messung. Kommunikation und Konfliktregelung stehen gleichauf in der Wichtigkeit allerdings ist ihr Abstand zur Bedeutung des Fachwissens nur noch halb so groß wie am Anfang.

Abb. 8: Systematischer Paarvergleich in der Trainingsgruppe.



Befund: Auch in der Trainingsgruppe sinkt die Wichtigkeitseinschätzung von Kommunikation, Konfliktregelung und Klassenführung - bleibt aber noch im „wichtigen“ Bereich. Die Klassenführung sinkt im Paarvergleich am stärksten.

Es wurde zum Zeitpunkt t3 auch geprüft, ob Geschlechtsunterschiede, Altersunterschiede und Unterschiede gemäß der fachlichen Orientierung geisteswissenschaftlich, sprachwissenschaftlich oder gesellschaftswissenschaftlich existieren. Hier konnten keine signifikanten Zusammenhänge gefunden werden.

3.3 Multiple Regressionen zur Abschätzung kausaler Beziehungen in Versuchs-, Kontroll- und Trainingsgruppe

Um die folgenden Auswertungen richtig bewerten zu können, sind einige Worte zur Erklärung der Multiplen Regression notwendig. Eine Multiple Regression entspricht einer multiplen Korrelation, das heißt einer Korrelation von zwei oder mehr Werten mit einer abhängigen Variablen. Der Buchstabe R bezeichnet eine solche Multiple

Korrelation im Unterschied zum Buchstaben r, der eine bivariate Korrelation bezeichnet. Multiple und bivariate Korrelation können ähnlich interpretiert werden: die maximal positive Korrelation ist in beiden Fällen 1,0 (ein Wert, der nie erreicht wird).

Bei der Multiplen Regression kann über die standardisierten Betagewichte der Beitrag der einzelnen unabhängigen Variablen zur abhängigen Variablen bestimmt werden. Je höher diese Werte sind, umso wichtiger ist deren Einfluss auf die Höhe der Multiplen Korrelation.

Ausbildung, Wissen und Wichtigkeit für das jetzige Verhalten

In den Tabellen 10 und 11 sind die Variablen Ausbildung, Wissen und Wichtigkeit von jeweils Kommunikation, Konfliktregelung und Klassenführung in ihrem Beitrag zum **jetzigen** Verhalten dargestellt. Und zwar für die erste und dritte Messung.

Beziehungsweise in der Kontrollgruppe für die erste (1) und die letzte Messung.

Tabelle 10: Multiple Regression (linear) der Einschätzung von Ausbildung, Wissen und Wichtigkeit auf die Einschätzung des jetzigen Verhaltens in der **Versuchsgruppe** - erste und vierte Messung. (*= signifikant 5%Niveau)

abhängige V	Multiple Korr.		Ausbildung		Wissen		Wichtigkeit	
	R1	R4	t1	t4	t1	t4	t1	t4
Jetziges Verhalten in...								
Kommunikation	0,44*	0,57*	0,16	-0,34*	0,34*	0,74*	0,02	-0,01
Konfliktregelung	0,67*	0,72*	-0,05	-0,03	0,70*	0,73*	0,15	0,19
Klassenführung	0,74*	0,63*	-0,17	0,09	0,81*	0,57*	0,11	-0,06

Es ist zunächst einmal erkenntlich, dass alle multiplen Korrelationskoeffizienten signifikant sind, das heißt: Ausbildung, Wissen und Wichtigkeit tragen signifikant zum jetzigen Verhalten bei. Des weiteren fällt auf, das insbesondere das Wissen signifikant zum jetzigen Verhalten beiträgt und eher weniger die Ausbildung. Ob die Referendare hier annehmen - ganz im Sinne der Selbstbildung - nur das, was in ihrer kognitiven Struktur angekommen ist, verändert das Verhalten oder, ob es generell

wichtiger ist, Wissen zu vermitteln als „auszubilden“, kann hier nicht entschieden werden.

Tabelle 11: Multiple Regression (linear) der Einschätzung von Ausbildung, Wissen und Wichtigkeit auf die Einschätzung des jeweiligen Verhaltens in der **Kontrollgruppe** - erste und letzte Messung (t1 u. t4) (*=signifikant 5% Niveau)

abhängige V	Multiple Korr.		Ausbildung		Wissen		Wichtigkeit	
	R1	R4	t1	t4	t1	t4	t1	t4
Jetziges Verhalten in...								
Kommunikation	0,51*	0,75*	0,10	0,14	0,44*	0,65*	0,00	0,10
Konfliktregelung	0,59*	0,46*	-0,47*	-0,12	0,77*	0,52*	-0,21	-0,10
Klassenführung	0,82*	0,55*	0,02	-0,15	0,78*	0,54*	0,09	0,21

Die neuen Pisa E Untersuchungen zum naturwissenschaftlichen Unterricht (2008) haben tatsächlich herausgefunden, dass ein handlungsorientierter Unterricht in den Naturwissenschaften keinesfalls immer von den besten Erfolgen begleitet wird, sondern dass ein kognitiv anspruchsvoller Unterricht deutlich bessere Resultate zeigt. Sollte das ein Hinweis darauf sein, dass Wissensvermittlung doch besser ist als Ausbildung? Jedenfalls ist bemerkenswert, dass dem jetzigen Wissenstand mehr Gewicht für das jetzige Verhalten beigemessen wird als der Ausbildung.

Befund: In Versuchsgruppe und Kontrollgruppe hat das jetzige Wissen auf das jetzige Verhalten den stärksten Einfluß - stärker als die Einschätzung der jetzigen Ausbildung. Keinen Einfluß hat die Wichtigkeitseinschätzung (multiple Regression).

Jetziges Verhalten als Ursache für erwartete Schwierigkeiten

In den folgenden zwei Tabellen (12 und 13) wird für Versuchs- und Kontrollgruppe das Verhalten in Kommunikation, Konfliktregelung und Klassenführungsverhalten am Ende der Ausbildung mit den Erwartungen von Schwierigkeiten in der Praxis in Verbindung gebracht. Das geschieht mit dem Ziel, den Beitrag der Ausbildung für die

subjektive Verringerung der Schwierigkeitserwartungen zu prüfen. Wer sich sicher fühlt und glaubt, Probleme meistern zu können - bewertet die zu erwartenden Schwierigkeiten des Alltags geringer.

Die Auswahl der vier Variablen erwarteter Schwierigkeiten geschahen im Anschluss an eine Faktorenanalyse der insgesamt sieben Variablen. Dabei wurden zwei Faktoren ermittelt: einer ist nur mit individueller Gerechtigkeit und Heterogenität zu bezeichnen und der andere enthielt die übrigen Schwierigkeiten bis auf die Kommunikation und die Verständlichkeit des Stoffes (Ladungen über. 60). Zur Markierung der beiden Faktoren und der Kommunikation wurden dann die vier in den Tabellen dargestellten Variablen ausgewählt.

Tabelle 12: Multiple Regression (linear) der Einschätzung des jetzigen Kommunikations-, Konfliktregelungs- und Klassenführungsverhaltens auf ausgewählte Variablen der zukünftigen Schwierigkeitseinschätzungen in der **Versuchsgruppe** (letzte Messung t4)(*=signifikant 5%Niveau)

erwartete Schwierigkeiten	R	jetziges Verhalten in...		
		Kommunikation	Konfliktregelung	Klassenführungsverhalten
ruhig und disziplinierter Unterricht	0,35	0,09	0,25	0,09
richtigen Ton treffen	0,41*	-0,04	0,30(*)	0,19
Konflikte schnell regeln	0,47*	0,01	0,51*	-0,22
allen gerecht werden	0,26	0,26	-0,08	0,06

Man erkennt in der Versuchsgruppe, dass nur im Bereich der Kommunikation (den richtigen Ton treffen) und der Konfliktregelung ein signifikanter (beziehungsweise bei der Kommunikation fast signifikanter) Beitrag der jetzigen Einschätzung von Konfliktregelungsverhalten ermittelt wird. Das kann auf die besondere Bedeutung des Konfliktregelungsverhaltens in der Versuchsgruppe zurückzuführen sein - hier

wurde ein spezifischer externer Kurs dazu angeboten. Weder das jetzige Kommunikationsverhalten noch das jetzige Klassenführungsverhalten stehen in einer systematischen Beziehung zur Erwartung oder Nichterwartung von Schwierigkeiten.

Tabelle 13: Multiple Regression (linear) der Einschätzung des jetzigen Kommunikations-, Konfliktregelungs- und Klassenführungsverhaltens auf ausgewählte Variablen der zukünftigen Schwierigkeitseinschätzungen in der **Kontrollgruppe** (letzte gemeinsame Messung t4)(*=signifikant 5%Niveau)

erwartete Schwierigkeiten	R	jetziges Verhalten in...		
		Kommunikation	Konfliktregelung	Klassenführungsverhalten
ruhig und disziplinierter Unterricht	0,51*	-0,32*	0,48*	0,00
richtigen Ton treffen	0,55*	0,42*	0,27	0,00
Konflikte schnell regeln	0,61*	-0,21	0,61*	0,03
allen gerecht werden	0,41(*)	0,34*	0,22	-0,21

In der Kontrollgruppe sieht das Bild anders aus. Hier trägt das jetzige Verhalten in Kommunikation und Konfliktregelung zur Wahrnehmung der Schwierigkeiten signifikant bei. Lediglich die Beurteilung der jetzigen Kommunikation verfehlt knapp den signifikanten Beitrag zur Konfliktlösung.

Befund: Sowohl in der Kontroll- wie auch Versuchsgruppe gibt es einen signifikanten Zusammenhang zwischen den erwarteten Schwierigkeiten „Konflikte schnell und effizient regeln zu können“ und der Selbsteinschätzung des jetzigen „Konfliktregelungsverhaltens“. Das jetzige Klassenführungsverhalten steht in keiner systematischen Beziehungen mit den erwarteten Schwierigkeiten. In der Versuchsgruppe gilt dies auch für die Kommunikation - allerdings nicht in der Kontrollgruppe.

Hängt die Verhaltenseinschätzung am Ende der Referendarzeit mit dem Uni-Wissen zusammen?

Als letzte Auswertung wurde die Ausgangseinschätzung von Ausbildung, Wissen und Wichtigkeit, die sich auf den universitärer vermittelten Stand richtete, in Verbindung gebracht mit dem Verhalten im Kommunikation, Konfliktregelung und Klassenführung am Ende der Referendarzeit. Damit sollte geprüft werden, ob es Nachwirkungen des universitär vermittelten Wissens bis ans Ende der Referendarzeit gibt.

Tabelle 14: Multiple Regression (linear) der Ausgangswerte (nach Uni) in Ausbildung, Wissen und Wichtigkeit auf die Einschätzung des Verhaltens der drei Themenbereiche Kommunikation, Konfliktregelung und Klassenführungsverhalten in der **Versuchsgruppe**. (*=signifikant 5%Niveau)

jetziges Verhalten in...	R	Ausgangseinschätzung nach Uni-Zeit in		
		Ausbildung	Wissen	Wichtigkeit
Kommunikation	0,45*	-0,02	0,47*	0,03
Konfliktregelung	0,51*	-0,33*	0,64*	0,13
Klassenführung	0,48*	-0,08	0,47*	-0,16

In der Versuchsgruppe ist überraschend, dass das universitäre Wissen einen signifikanten Beitrag zu allen drei Bereichen leistet. Die multiplen Korrelationskoeffizienten unterscheiden sich in ihrer Höhe kaum von denen zwischen den unabhängigen und abhängigen Variablen am Ende der Referendarzeit.

Tabelle 15: Multiple Regression (linear) der Ausgangswerte (nach Uni) in Ausbildung, Wissen und Wichtigkeit auf die Einschätzung des Verhaltens der drei Themenbereiche Kommunikation, Konfliktregelung und Klassenführungsverhalten in der **Trainingsgruppe der Versuchsgruppe**. (*=signifikant 5% Niveau)

jetziges Verhalten in...	R	Ausgangseinschätzung nach Uni-Zeit in-		
		Ausbildung	Wissen	Wichtigkeit
Kommunikation	0,39	0,31	0,15	0,06
Konfliktregelung	0,19	-0,01	0,20	0,02
Klassenführung	0,44	0,17	-0,02	0,44

Überraschenderweise lässt sich dieser Zusammenhang in der Trainingsgruppe der Versuchsgruppe nicht herstellen. Das kann aber mit der geringen Personenzahl in der Trainingsgruppe zusammenhängen. Die absoluten Werte der multiplen Korrelation und auch der standardisierten Beta-Gewichte sind allerdings außerordentlich gering.

Daraus kann man folgendes schließen: über Wissen wird das Verhalten der Referendare verändert. Aus diesem Grunde kann auch das universitär erworbene Wissen noch das Verhalten am Ende der Referendarzeit beeinflussen. Zumal dieses Wissen in ähnlicher Form auch von den Fachleitern vertreten wird. Deshalb gibt es eine Kontinuität der Wissensvermittlung zwischen Universität und zweiter Ausbildungsphase.

Befund: Auch das Anfangswissen, das aus der universitären Ausbildung stammt, hat einen Einfluss auf die Einschätzung des „jetzigen Verhaltens“ in Kommunikation, Konfliktregelung und Klassenführung am Ende des Projektzeitraums.

5. Diskussion

Zunächst einmal muss relativierend angemerkt werden, dass in der Trainingsgruppe längsschnittlich nur eine relativ geringe Anzahl von Referendaren enthalten war (N=18 bis N=22). Das hat statistisch den Effekt, dass nur erhebliche Unterschiede signifikant werden. Auch ist die Signifikanzprüfung nicht unbedingt immer mit Relevanz zu übersetzen. Relevant können auch kleinere und nicht signifikante Verbesserungen sein.

Zunächst konnte aber gesichert werden, dass die Anfangsunterschiede zwischen Trainingsgruppe und Kontrollgruppe nicht signifikant voneinander abweichen. Bis auf zwei Items im Bereich der Beurteilung von Kommunikations- und Klassenführungsverhalten gibt es keine signifikanten bzw. relevanten Unterschiede zwischen Trainings- und Kontrollgruppe. Das gilt auch für die Beurteilung der praktischen Herausforderungen unterschiedlicher Lehraufgaben, die alle nicht signifikant sind. Mit den entsprechenden Diskriminanzanalysen konnte dieser Befund erhärtet werden. Es ist aus unserer Sicht zu beurteilen: Im Wesentlichen sind Trainings- und Kontrollgruppe zu Beginn des Experimentes vergleichbar. Das ist eine Grundvoraussetzung für die Ermittlung von tatsächlichen Effekten.

Im Folgenden werden die elf wesentlichen Befunde, die auch schon in der Zusammenfassung und im Text genannt wurden kommentiert und diskutiert.

Befund 1: Die Kontrollgruppe bewertet am Ende des Versuchs die Ausbildung, das Wissen und das jetzige Verhalten subjektiv geringfügig besser.

Der Befund ist zwiespältig zu bewerten: eine bessere Beurteilung kann auch darauf zurückzuführen sein, dass die Ausbildung insgesamt adressatennah war - im Klartext: sie hat nicht dazu geführt, dass die eigene Ausbildung kritischer und anspruchsvoller bewertet wird. Die professionellen Trainingselemente können auch dazu führen, dass „man weiß, dass man nichts weiß“. Gemessen an optimalen Formen der Realisierung erscheint das, was man als Referendar beherrscht nunmehr als „nicht ganz so toll“. Auf der anderen Seite ist der Befund aber ein ernst zu nehmender Hinweis darauf, dass die Kontrollgruppe auch eine ordentliche

Ausbildung abgeliefert hat. Der Befund 1 ist aber nur im Zusammenhang mit Befund 2 zu interpretieren.

Befund 2: Die Versuchsgruppe bewertet in zentralen Bereichen die kommenden Herausforderungen der Praxis (Disziplin, Konfliktlösung und Stoff verständlich machen) als weniger schwierig - obwohl die Ausgangslage in beiden Gruppen identisch war.

Hier zeigt sich möglicherweise doch ein Vorteil der Versuchsgruppe, der auch den Befund 1 in einem anderen Licht erscheinen lässt. Die Versuchsgruppe schreckt der Lehrer-Alltag weniger - offenbar gewinnen sie durch das Training auch Kompetenz (oder besser: Kompetenzvermutungen), dass man die beschriebenen Probleme des Schulalltags mit dem Gelernten gut meistern kann.

Befund 3: Mit zunehmender Praxis und Ausbildung wird die Wahrnehmung der Wichtigkeit der drei Bereiche Kommunikation, Konfliktregelung und Klassenführung schwächer. Die Unterschiede in der Beurteilungsskala sind nicht sehr dramatisch. Es handelt sich um Unterschiede, die zwischen 0,2 und 0,6 Skalenstufen schwanken.

Wer von der Uni kommt, hält sich mit Recht für zu wenig praktisch ausgebildet. Folglich erscheint alles wichtig, für das man in der Uni keine Sicherheit erlangt hat: die zwischenmenschliche Kommunikation, das Verhalten in Konflikten und der Umgang mit einer großen und unruhigen Klasse. Folglich hält man die zukünftige Ausbildung in diesen Bereichen (der Praxis) für besonders wichtig. Mit der zunehmenden eigenen Praxis verändert sich aber die Wichtigkeit: der Trend geht in Richtung „weniger wichtig“. Befund 4 gehört aber noch dazu.

Befund 4: Mit Hilfe einer besonders zuverlässigen Skalierungsmethode (Paarvergleich) kann nicht nur der o.g. Befund bestätigt werden, dass die subjektiv empfundene Wichtigkeit der ausgewählten Ausbildungsbereiche während des Referendariats sinkt, sondern auch dass die Wichtigkeitsbeurteilung aller Variablen näher zusammenrückt. Die eher psychologischen Bereiche Konfliktregelung, Kommunikation und Klassenführung rücken näher an die Beurteilung der fachdidaktischen Bereiche Unterrichtsablauf und Fachwissen heran. Besonders sinkt die Wichtigkeitsbeurteilung der Klassenführung in Versuchs- und Kontrollgruppe.

Aber: Fachwissen und Unterrichtsablauf bleiben weniger wichtig - alle anderen Bereiche erscheinen wichtiger. Nur ist der Abstand an Wichtigkeit geschrumpft. Der starke Abstieg des Klassenführungsverhaltens veranlasst zu Fragen an die Ausbildung. Ist der Zusammenhang zwischen Kommunikation, Konfliktregelung und Klassenführung deutlich genug gemacht worden? Wurde deutlich, dass Klassenführung zur wirksamen Prävention von Konflikten führt? Konnte verhindert werden, dass der Eindruck entsteht, wenn man die Regelung von Konflikten beherrscht, dann braucht man was anderes nicht mehr zu lernen?

Befund 5: Bezogen auf Ausbildung, Wissen und Verhalten zeigen die Ergebnisse der Trainingsgruppe innerhalb der Versuchsgruppe über die vier Messzeitpunkte hinweg eine Verbesserung der subjektiven Einschätzung in allen drei Bereichen (Kommunikation, Konfliktregelung und Klassenführung). Der Verlauf ist nicht in allen Fällen kontinuierlich.

Die Trainingsgruppe - gewissermaßen der Kern der Versuchsgruppe - zeigt in allen Variablen einen positiven Verlauf über die vier Messzeitpunkte. Das ist ein Erfolg des Bielefelder Projektes - der allerdings leider nicht auf die versuchsgruppeninterne Kontrollgruppe übertragbar war (siehe Befund 6). Dazu müssten mehr Ressourcen her.

Befund 6: Die Trainingsgruppe innerhalb der Versuchsgruppe ist in den meisten Variablen besser als die versuchsgruppeninterne Kontrollgruppe. Das muss bei der Bewertung der Ergebnisunterschiede zwischen Kontrollgruppe und Versuchsgruppe berücksichtigt werden.

Ebenfalls ein deutlicher Erfolg: eigentlich hätte nur die Trainingsgruppe den Namen „Versuchsgruppe“ verdient gehabt. Dennoch: in einigen Variablen unterscheidet sich die versuchsgruppeninterne Kontrollgruppe nicht und in einigen wenigen ist sie sogar leicht besser. Anlass zum Nachdenken also.

Befund 7: Referendare in der Trainingsgruppe bewerten die Schwierigkeiten, Schüler ruhig und diszipliniert zu unterrichten, ebenso wie allen Schülern gerecht

werden und den Umgang mit Heterogenität unverändert schwierig. Die Ausbildung hat auf die Erwartungsänderung keinen Einfluss.

Eine zentrale Forderung der bildungspolitischen Phraseologie ist die individuelle Förderung. Das zu fordern und darüber zu reden ist billig - es machen ist schwierig, wenn nicht gar unmöglich. Diesen Nichtbefund muss man gelassen hinnehmen. Das Schulsystem ist mit seinen aktuellen Ressourcen nicht wirklich zur individuellen Förderung geeignet - in keiner Schulstruktur und in keinem Land.

Befund 8: Auch in der Trainingsgruppe sinkt die Wichtigkeitseinschätzung von Kommunikation, Konfliktregelung und Klassenführung - bleibt aber noch im „wichtigen“ Bereich. Die Klassenführung sinkt im Paarvergleich am stärksten.

Hierzu ist nichts weiter zu sagen als das, was zu Befund 4 schon ausgeführt wurde. Auch in der Trainingsgruppe ist derselbe Mechanismus wirksam.

Befund 9: In Versuchsgruppe und Kontrollgruppe hat das jetzige Wissen auf das jetzige Verhalten den stärksten Einfluss - stärker als die Einschätzung der jetzigen Ausbildung. Keinen Einfluss hat die Wichtigkeitseinschätzung (multiple Regression).

Merkwürdig: die Bewertung der jetzigen Ausbildung wirkt kaum auf die Einschätzung des jetzigen Verhaltens. Wohl aber das Wissen. Die Ausbildung hat das Wissen vermittelt - darauf kommt es wohl an. Der Befund passt zu aktuellen Befunden der Unterrichtsforschung (PISA E 2008), dass der entscheidende Mechanismus nicht der Prozess, sondern der erlebte Gewinn an Einsicht ist.

Befund 10: Sowohl in der Kontroll- wie auch Versuchsgruppe gibt es einen signifikanten Zusammenhang zwischen den erwarteten Schwierigkeiten „Konflikte schnell und effizient regeln zu können“ und der Selbsteinschätzung des jetzigen „Konfliktregelungsverhaltens“. Das jetzige Klassenführungsverhalten steht in keiner systematischen Beziehungen mit den erwarteten Schwierigkeiten. In der Versuchsgruppe gilt dies auch für die Kommunikation - allerdings nicht in der Kontrollgruppe.

Ein Erfolg für die Referendarausbildung: das Zutrauen in die Konfliktregelungsfähigkeit ist gestiegen. Wenn man etwas kann (Konflikte regeln) wird man dessen Schwierigkeiten als nicht unüberwindlich einschätzen. Enttäuschung für das Klassenführungsverhalten: offenbar steht es in keiner Beziehung zur Lösung praktischer Probleme. Das würde aber der empirischen Forschungslage widersprechen - dessen Beitrag ist objektiv nachgewiesen und mehrfach repliziert. Also: Anregung zum Nachdenken, ob das Klassenführungsverhalten in der richtigen Form, Art und Weise vermittelt wurde.

Befund 11: Auch das Anfangswissen, das aus der universitären Ausbildung stammt, hat einen Einfluss auf die Einschätzung des „jetzigen Verhaltens“ in Kommunikation, Konfliktregelung und Klassenführung am Ende des Projektzeitraums.

Erfolg für die Uni-Ausbildung: das Anfangswissen hinterlässt seine Spuren in der Verhaltenseinschätzung am Ende der (Referendars-) Projektzeit. Natürlich ist das Wissen, das im Referendariat vermittelt wird, nicht grundverschieden von dem, was in der Uni vermittelt wurde. Wenn Praxis und Wissenschaft genügend integriert sind...

Literatur:

Dollase, R. (1995). Die virtuelle oder psychologische Reduzierung der Schulklassengröße. *Bildung und Erziehung*, 48(2), 131 - 144.

Eichhorn, C. (2008). *Classroom - Management*. Stuttgart: Klett Cotta.

Evertson, C. M., & Weinstein, E. A. (Eds.). (2006). *Handbook of Classroom Management: Lawrence Erlbaum Associates*.

Lösel, F., & Nowack (1987). Evaluationsforschung. In J. Schultz - Gambard (Ed.), *Angewandte Sozialpsychologie* (pp. 57 - 90). Weinheim: PVU.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge Base for School Learning. *Review of Educational Research*, 63(3), 249 - 294.

ANHANG

UNIVERSITÄT BIELEFELD

Wissenschaftliche Begleitung TIPP-Projekt – Evaluation – Prof. Dr. Rainer Dollase und
Dipl.Päd. Odette Selders

FRAGEBOGEN EU TIPP PROJEKT

Liebe Referendarinnen und Referendare,

Evaluation (Bewertung) muss sein, die Zuweisung von Projektmitteln ist daran gebunden – deshalb bitten wir Sie herzlichst an der folgenden kurzen Befragung teilzunehmen, die Sie ja schon von der Anfangserhebung her kennen.

Antworten Sie spontan und zügig. Die Befragung ist anonym – eine Identifizierung Ihres Fragebogens ist allerdings deshalb nötig, damit der 2. Fragebogen Ihrem ersten zugeordnet werden kann, um damit Veränderungen zu erfassen.

Die Codierung ist allerdings unverfänglich:

Geburtsmonat Ihrer Mutter (zweistellig):

Zweiter Buchstabe im Vornamen Ihres Vaters (einstellig):.....

Letzter Buchstabe Geburtsort.....

Und nun bitte ausfüllen...

Herzliche Grüße

Rainer Dollase und Odette Selders

1. Es geht um drei Themen: Kommunikation – Konfliktbearbeitung im Unterricht – Klassenführung (classroom management). Beurteilen Sie bitte in **Schulnoten von 1 = sehr gut bis 6 = ungenügend** die folgenden Bereiche.

Die Ausbildung in Kommunikation im Referendariat. Note:	Die Ausbildung in Konfliktbearbeitung im Referendariat. Note:	Die Ausbildung in Klassenführung im Referendariat. Note:
Mein heutiges Wissen über Kommunikation Note:	Mein heutiges Wissen über Konfliktbearbeitung Note:	Mein heutiges Wissen über Klassenführung Note:
Mein jetziges Kommunikationsverhalten Note:	Mein jetziges Verhalten in Konflikten Note:	Mein jetziges Klassenführungsverhalten Note:

2. Wie beurteilen Sie die Wichtigkeit der drei Bereiche Kommunikation – Konfliktbearbeitung – Klassenführung ? Geben Sie wieder Schulnoten von **1 = sehr wichtig bis 6 = sehr unwichtig**.

Wichtigkeit einer Verbesserung der Kommunikation Note:	Wichtigkeit einer Verbesserung der Konfliktbearbeitung Note:	Wichtigkeit einer Verbesserung der Klassenführung Note:
---	---	--

3. Entscheiden Sie sich jeweils für eine von zwei Alternativen!

Was ist für die Praxis wichtiger? (Bitte jeweils je Zeile eine Alternative ankreuzen)

Fachwissen in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Unterrichtsabläufe in meinen Fächern
Fachwissen in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Kommunikation
Fachwissen in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Konfliktbearbeitung
Fachwissen in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Klassenführung
Unterrichtsabläufe in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Kommunikation
Unterrichtsabläufe in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Konfliktbearbeitung
Unterrichtsabläufe in meinen Fächern	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Klassenführung
Verbesserung meiner Kommunikation	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Konfliktbearbeitung
Verbesserung meiner Kommunikation	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Klassenführung
Verbesserung meiner Konfliktbearbeitung	<input type="checkbox"/>	oder	<input type="checkbox"/>	Verbesserung meiner Klassenführung

3. Wo erwarten Sie in Ihrer zukünftigen Praxis die größten Herausforderungen? Geben Sie für die Höhe Ihrer Schwierigkeitserwartungen wieder Schulnoten von **1= geringe Schwierigkeiten bis 6 = allergrößte Schwierigkeiten**.

Die Schüler/innen im Unterricht ruhig und diszipliniert zu halten. Note Schwierigkeit:	Das nötige Fachwissen sicher zu beherrschen. Note Schwierigkeit:	Den richtigen Ton im Umgang mit den Schülern/innen zu treffen. Note Schwierigkeit:
Konflikte mit Schülern/innen schnell und effektiv lösen. Note Schwierigkeit:	Den Schülern/innen den Stoff verständlich vermitteln können. Note Schwierigkeit:	Allen Schülern gerecht zu werden (Individualisierung) Note Schwierigkeit:
Der Umgang mit Heterogenität der Schülerschaft Note Schwierigkeit:		

Zum Schluss: Geschlecht? (m/w)..... Alter:.....Jahre

Wie beurteilen Sie Ihre fachlichen Interessen?

eher naturwissenschaftlich eher gesellschaftswissenschaftlich eher sprachwissenschaftlich

Das wär's für heute – vielen Dank!