

Design, Durchführung und Bewertung der Tests

von: *Rainer Hampel*

1. Testdesign

Das zu Beginn des COMCULT-Projekts geplante Testdesign (siehe Netzwerkinformationen => Netzwerkbeschreibung, Punkt 11 Evaluation) musste, nachdem die Arbeit in den Partnerschulen begonnen hatte, leicht modifiziert werden.

A) Faktenwissen – Werteerziehung - Wertebewusstsein

Faktenwissen über einen für alle Schüler verbindlichen Fragekatalog abzufragen, erwies sich zwar als durchführbar, überlastete aber den Fragebogen zeitlich zu sehr, ohne einen entsprechenden Aussagewert zu haben. Außerdem wäre es den Lehrkräften zwischen Pre- und Posttest möglich gewesen, ihre SchülerInnen gezielt auf die richtigen Antworten vorzubereiten.

Deswegen wurden bestimmten Werten eine Reihe von Behauptungen/Meinungen (statements) zugeordnet und die Jugendlichen gebeten, ihre Zustimmung oder Ablehnung durch Ankreuzen der entsprechenden Spalten auszudrücken. Auf diese Weise war es möglich, den Wissensstand der SchülerInnen teilweise indirekt über ihre Meinungen zu erfahren.

Alle Werte im Kapitel A) „Wertebewusstsein“ des Fragebogens wurden dem „Vertrag über eine Verfassung für Europa“ Teil I, Artikel 1 und 2, entnommen, da sie für das Zusammenleben in der EU von tragender Bedeutung sind:

- | | |
|-----------------------|------------------|
| - Freiheit | - Toleranz |
| - Demokratie | - Gerechtigkeit |
| - Gleichheit | - Solidarität |
| - Rechtsstaatlichkeit | - Menschenrechte |
| - Pluralismus | - Heimat, Nation |

In der **Urfassung des Fragebogens** kann man sehen, dass diesen Werten jeweils sieben Behauptungen/Meinungen zugeordnet wurden, um ein differenziertes Bild über die Kenntnisse und das Wertebewusstsein der SchülerInnen zu erhalten. Diese Ordnung wurde auch für die Darstellung der Ergebnisse (Basisdaten) beibehalten.

Für die **endgültige Version** des Fragebogens und aller Übersetzungen wurden die Behauptungen/Meinungen im Kapitel A), C) und teilweise auch in D) gemischt, um die Jugendlichen zu zwingen, immer wieder spontan ihre Meinung zu sagen und ihnen keine Zeit zu lassen, ihre Kreuze nach anderen Gesichtspunkten bewusst zu ordnen.

B) Wertevermittlung

In diesem Kapitel sollte geklärt werden, welche Werte durch welche Gruppe oder Einrichtung am besten vermittelt werden können. Da bestimmte Werte durchaus in unterschiedlichen Gruppen bzw. Einrichtungen eine gewichtige Rolle spielen, konnten die SchülerInnen hier maximal zwei Kreuze pro Wert setzen. In den meisten Fällen haben sie das auch gemacht, aber es kommt auch manchmal vor, dass nur eine Spalte pro Wert angekreuzt wurde.

C) Verhaltensdisposition

In der Erfassung des Fragebogens und in den Ergebnissen (Basisdaten) sind die folgenden Einstellungen oder Charaktereigenschaften aufgelistet:

- | | |
|--------------------|---------------|
| - Selbstbestimmung | - Sicherheit |
| - Universalismus | - Macht |
| - Wohlwollen | - Leistung |
| - Konformität | - Hedonismus |
| - Tradition | - Stimulation |

Diesen einzelnen Verhaltensdispositionen wurden jeweils zwei Aussagen zugeordnet, auf die die Jugendlichen entweder mit Zustimmung oder Ablehnung reagieren konnten.¹

D) Unterrichtsmethoden und Medienkompetenz

Am Anfang der Arbeit im COMCULT-Netzwerk stand die Frage, ob bestimmte Unterrichtsmethoden unterschiedliche Wirkung haben auf:

- die Nachhaltigkeit der Kenntnisvermittlung
- das Wertebewusstsein oder
- auf Verhaltensdispositionen.

Leider konnte das Testdesign auf diese Frage keine zweifelsfreien Antworten geben, da sich schon bei der Erprobung der ersten Fragebogenentwürfe im Jahre 2004 zeigte, dass sich keine Partnerschule finden ließ, an der bestimmte Klassen oder Gruppen nur nach einer Methode allein unterrichtet wurden. In den meisten Fällen kombinierten die Lehrkräfte verschiedene Lehr- und Lernmethoden miteinander, sodass es nicht möglich war, spezifische Unterrichtswirkungen sicher auf die eine oder andere Methode zurückzuführen.

Auf der nächsten Seite findet sich eine Tabelle mit

- der Unterrichtsorganisation
- den behandelten Themen und
- den dabei angewandten Unterrichtsmethoden.

Die Angaben über die Anzahl der Lehrkräfte und der SchülerInnen pro Klasse oder Gruppe beziehen sich auf die gesamte Projektlaufzeit von drei Jahren, d.h. auf eine viel größere Anzahl von Jugendlichen als die, welche im dritten Projektjahr an der Fragebogenaktion teilnahm.

Um in der gegebenen Lage doch noch zu Informationen über bestimmte Lehr- und Lernmethoden zu gelangen, wurden die SchülerInnen gebeten, direkt ihre Bewertung wiederzugeben, und zwar über:

- den Frontalunterricht
- den Gruppenunterricht und über
- die Nutzung des Internets für Unterrichtszwecke.

Fünf Argumente für den Frontalunterricht und fünf für den Gruppenunterricht konnten von den SchülernInnen mit bis zu 3 Kreuzen gewichtet werden. Aus der Addition der Kreuze pro Argument lässt sich ablesen, welche Aspekte der einen oder anderen Methode in den Augen der Jugendlichen besonders geschätzt werden.

¹ Einzelne Aussagen (Items) wurden aus folgendem Artikel entnommen, modifiziert und zum Teil ergänzt: JULIA ISER, PETER SCHMIDT. Gefährliche Werte? in „Deutsche Zustände“, Folge 2, Hrsg. WILHELM HEITMEYER, Suhrkamp Verlag, Frankfurt a.M. 2003, S. 61-77

Die Unterrichtsorganisation in den einzelnen Schulen des COMCULT Netzwerks

		DE 1	GR 2	ES 3	IT 4	NL 5	UK 6	BG 7	CY 8	HU 9	PL 10
3.1	Frontalunterricht Lehrkräfte insgesamt:	10 L * K mit 29 S G mit 9 S	1 L * K mit 22 S	9 L * K mit 25 S		6 L * K mit 23 S G mit 7 S	2 L	2 L * K mit 26 S G mit 13 S	10 L K mit 18 S G mit 4 S	9 L * K mit 32 S G mit 16 S	6 L K mit 30 S G mit 10 S
	Thema 1	2 K und 5 G	1 K	2 K		1 K und 2 G		2 K	2 K	3 K und 6 G	1 K und 3 G
	Thema 2	1 K		4 K		1 K und 2 G	2 K mit 16S	2 K	*	1 K und 2 G	1 K
	Thema 3					1 K und 2 G		2 K	*	4 K und 6 G	
	Thema 4		1 K			1 K		2 K	2 K		
3.2	Gruppenunterricht mit einer Lehrkraft ohne zusätzliches Internetmaterial	*	*	4 L * 4 K G je 5 S	3 L K mit 13 S G je 3 S	4 L * K mit 23 S G mit 5 S		2 L		3 L K mit 32 S G mit 16 S	
	Thema 1	1 K und 5 G	1 K mit 4 G	10 G	4 G	4 G		2 G	2 K	3 K und 6 G	
	Thema 2	1 K		15 G	1 G	4 G		2 G		1 K und 1 G	1 K und 3 G
	Thema 3					4 G	4 G mit 6 S	2 G		2 K und 4 G	1 K
	Thema 4		1 K mit 4 G			4 G			2 K		
3.3	Fächerübergreifender Unterricht mit mehreren Lehrkräften ohne zusätzliches Internetmaterial	4 L		4 L K mit 25 S	3 L K mit 13 S	6 L		3 L K mit 26 S G mit 13 S		4 L K mit 32 S G mit 16 S	2 L
	Thema 1					1 K		2 K		1 K	
	Thema 2			4 K	1 K	4 G		2 G		1 K	
	Thema 3	1 K			1 K	4 G				3 K	1 K
	Thema 4										
	Fächerübergreifender Unterricht mit mehreren Lehrkräften und mit zusätzlichem Internetmaterial			2 L K mit 25 S		6 L		3 L	10 L	3 L C mit 32 S G mit 16 S	
	Thema 1					1 K		2 G	2 K	1 K und 1 G	
	Thema 2			2 K		4 G		2 G			
	Thema 3					4 G				2 K und 2 G	
Thema 4								2 K			
3.4	Grenzüberschreitender (projektorientierter) Unterricht mit Hilfe der IKT	1 K mit 31 S	1 K mit 20 S	3 K mit durchschn. 25 S	3 K mit durchschn. 12 S	1 K mit 23 S	2 K mit durchschn. 12 S	2 G mit jeweils 13 S	1 K mit 18 S	2 K und 1 G mit 6 S	1 K mit 30 S
	mit wie viel Lehrkräften?	3 L	2 L	3 L	3 L	3 L	1 L	2 L	3 L	3 L	2 L

Ziffer und L = Anzahl der Lehrkräfte | Ziffer und K = Anzahl der Klassen | Ziffer und G = Anzahl der Gruppen | Ziffer und S = Anzahl der Schüler pro Klasse oder Gruppe

* Lehrkräfte haben teilweise auch andere Unterrichtsmethoden benutzt.

IKT = Informations- und Kommunikationstechnik (hier: Gemeinsame Produktion eines Videofilms – teilweise via Internet)

In der Urfassung des Fragebogens kann man sehen, welche Behauptungen/Meinungen benutzt wurden, um zu erfahren, wie die SchülerInnen folgende Gesichtspunkte der Internetnutzung im Unterricht bewerten:

- Informationsumfang und –zugang
- Hinderungsgründe für eine Nutzung des Internets
- Vorteile einer Nutzung
- Nachteile einer Nutzung.

Die tatsächliche Nutzung des Internets im Laufe des Projekts lässt sich aus den Antworten zu D 64 entnehmen in Kombination mit den Angaben der Lehrkräfte zur Unterrichtsorganisation (Seite 3, Punkt 3.3 und 3.4).

E) Persönliche Einschätzungen

Am Ende des Fragebogens wurden noch eine Reihe von Behauptungen und Fragen eingefügt, um die Meinung der SchülerInnen zu erfahren über:

- Die Rolle der Politik in der Schule
- Einstellung zu ethischen Fragen
- Selbsteinschätzung im Unterricht
- Prioritäten in der nationalen Politik
- Heimat- und Nationalgefühl

Die zugehörigen Behauptungen/Meinungen und Fragen finden sich geordnet in der **Urfassung des Fragebogens**. Im tatsächlich **benutzten Fragebogen und allen Übersetzungen** sind die einzelnen Behauptungen/Meinungen zum Teil gemischt.

2. Die Durchführung der Tests

Bereits während des zweiten Projektjahres wurden Testentwürfe in wenigen Schulklassen in Deutschland, Ungarn und Polen erprobt, um zu erfahren, wie viel Zeit für seine Durchführung nötig ist und welche Einzelheiten (Items) des Fragebogens hinreichend aussagekräftig sind. Die dabei gewonnenen Erkenntnisse gingen in die endgültige Version der Fragebögen ein.

Ursprünglich war es Ziel der externen Evaluation, Unterrichtswirkungen zu messen. Deswegen wurde zu Beginn des Unterrichts in den Klassen, die für eine Mitarbeit im COMCULT-Netzwerk ausgewählt waren, der Fragebogen als **Pretest** ausgegeben, um die Anfangssituation dokumentieren zu können. Nach einigen Monaten des Unterrichts über bestimmte Werte mussten dieselben SchülerInnen den gleichen Fragebogen als **Posttest** nochmals ausfüllen. Aus 25 Klassen oder Gruppen kamen 461 Fragebögen zusammen. Allerdings fehlte bei etwa einem Viertel entweder der Pretest eines/r Schülers/Schülerin oder der entsprechende Posttest. So blieben am Ende **365** Pre- und Posttestpaare übrig von Jugendlichen, die beide Tests ausgefüllt hatten. Nur diese Fragebögen wurden in die Auswertung einbezogen, und alle anderen aussortiert. Die nächste Seite enthält eine Liste mit Angaben über die beteiligten Schulen, die Anzahl der SchülerInnen, die an der Fragebogenaktion mitgewirkt hatten, und auch die Zeiträume zwischen Pre- und Posttest.

3. Bewertung der Tests

3.1 Die Reliabilität

Nicht alle Kapitel des COMCUT-Fragebogens erfüllen hinsichtlich der Zuverlässigkeit oder Messgenauigkeit Anforderungen, die man von Testverfahren in den Sozialwissenschaften gewohnt ist.

Die Durchführung der Pre- und Posttests

The realization of pre- and posttests

Land Coun-try	Ort Place	Schule School	Pretest Datum Date	Posttest Datum Date	Differenz in Monaten Difference in months	Alter der Schüler Age of students	Schüler- zahl Number of students
DE	Freiburg	MAX-WEBER-SCHULE	23.09.2005	05.07.2006	9	Jahre/ Years 16-17-18	30
DE	Freiburg	MAX-WEBER-SCHULE	27.09.2005	11.07.2006	9	16-17-18	25
DE	Freiburg	MAX-WEBER-SCHULE	27.09.2005	11.07.2006	9	16-17	7
DE	Freiburg	MAX-WEBER-SCHULE	27.09.2005	18.07.2006	9	17-18-19	9
DE	Freiburg	MAX-WEBER-SCHULE	04.10.2005	25.07.2006	9	16-17-18	23
DE	Freiburg	MAX-WEBER-SCHULE	04.10.2005	04.07.2006	9	16-17-18	14
DE	Freiburg	MAX-WEBER-SCHULE	04.10.2005	11.07.2006	9	18-19-20	9
GR	Pallini Attikis	ELLINOGERMANIKI AGOGI	12.10.2005	04.05.2006	7	15-16	19
ES	Cazorla	I.E.S. CASTILLO DE LA YEDRA	18.10.2005 17.10.2005 18.10.2005	16.06.2006 13.06.2006 16.06.2006	8 8 8	16-17 16-17 16-17	13 17 20
IT	Lecce	I.T.S. G. DELEDDA	17.10.2005 10.10.2005	02.05.2006 20.04.2006	6 6	17-18 17-18	11 12
NL	Utrecht	ST. GREGORIUS COLLEGE	16.11.2005	12.05.2006	6	14-15	17
BG	Smoljan	TECHNICAL SCHOOL OF ECONOMICS	17.10.2005	27.04.2006	6	18	18
CY	Larnaca	AMERICAN ACADEMY LARNACA	15.09.2005	08.12.2005	3	17-18	15
HU	Veszprém	VESZPRÉMI KÖZGAZDASÁGI SZAKKÖZÉPISKOLA	14.10.2005	21.04.2006	6	15-16	30
PL	Bierun	LICEUM OGÓLNOKSZTAŁCĄC E im. Powślańców Śląskich	18.10.2005	20.04.2006	6	17-18	23
Gesamtzahl der SchülerInnen mit Pre- und Posttest Total number of students with Pre- and Posttest							312

Die folgenden Klassen * (Kontrollgruppen) haben keinen spezifischen Unterricht über Werte bekommen:
The following classes * (control groups) did not get any specific lessons about values:

CZ*	Vsetín	MASARYKOVA GYMNÁZIUM	13.12.2005 05.12.2005	06.09.2006 05.09.2006	9 9	15-16 17-18	(15) (17)
DE*	Hildesheim	BISCHÖFLICHES GYMNASIUM JOSEPHINUM	08.02.2006	30.06.2006	5	16-17	(21)
Gesamtzahl der SchülerInnen mit Pre- und Posttest (Kontrollgruppen) Total number of students with Pre- and Posttest (control groups)							(53)
Gesamtzahl der SchülerInnen mit Pre- und Posttest Total number of students with Pre- and Posttest							365

Zwar gibt es im Fragebogen Kapitel mit ausreichender Reliabilität, denn auch bei der Wiederholung der Befragung im Posttest zeigten sich ähnliche Resultate wie beim Pretest, so z.B. im Kapitel B) Wertevermittlung, C) Verhaltensdisposition oder D) Unterrichtsmethoden. Aber gerade im Kapitel A) Wertebewusstsein kam es ja darauf an zu erfahren, ob durch einen gezielten Unterricht das Wertebewusstsein der Jugendlichen messbar verändert werden kann.

Vermutlich würde auch eine Befragung in einem ganz anders zusammengesetzten Schulnetzwerk ebenfalls zu anderen Ergebnissen führen.

3.2 Die Validität

Wie die Ergebnisse in Kapitel A) zeigen, gibt es punktuell zahlreiche, gewiss nicht zufällige Veränderungen zwischen Pre- und Posttest. Es ist aber gar nicht sicher, ob sie alle auf einen Unterricht zurückzuführen sind, der im Sinne des COMCULT-Netzwerks gezielt aufklärend wirken wollte. Auch in den beiden Kontrollklassen – ohne spezifischen Unterricht über Werte – lassen sich eine Reihe von Veränderungen zwischen Pre- und Posttest beobachten.

Viel größer sind aber häufig die Unterschiede bei den Pre- und Posttests zwischen den einzelnen Schulen. Das deutet daraufhin, dass der Test besonders im Kapitel A) neben Unterrichtswirkungen auch noch andere Wirkungen misst, die aus Faktoren herrühren, welche durch das COMCULT- Netzwerk nur sehr begrenzt beeinflusst werden konnten. Deswegen besitzt der Test als Instrument der Messung von Unterrichtswirkungen nur eine begrenzte Validität.

3.3 Die Generalisierbarkeit der Testergebnisse?

Die bisherigen Überlegungen legen es nahe, bei der Generalisierung der Fragebogenergebnisse sehr vorsichtig zu sein. Die Bedingungen, unter denen die Daten erhoben wurden, waren z.T. sehr ungleich:

- Unterschiedliches Alter der SchülerInnen zwischen 14 und 19 Jahren
- Unterschiedliche Schultypen: Allgemeinbildende und berufliche Schulen, private und öffentliche Schulen
- Unterschiedliche Stundentafeln (Fächer)
- Unterschiedliche Lehrpläne
- Verschiedene Unterrichtsmethoden
- Unterschiedliche fachliche Voraussetzungen bei den Lehrkräften, z.T. begrenzt auf nur ein Fach, z.T. im Team mit unterschiedlichen Fachvertretern.

Ein besonderes Problem in Hinblick auf allgemeine Aussagen über Testergebnisse ist die verhältnismäßig kleine Anzahl von SchülerInnen, die pro Klasse bzw. pro Schule am Pre- und Posttest teilgenommen haben. Deswegen muss man sich bei jeder Interpretation auch der %-Werte immer bewusst sein, dass in einer kleinen Klasse von nur 18 SchülerInnen bereits drei ein prozentuales Gewicht von 17% haben. Rein statistisch kann man auf so minimale Veränderungen kaum gewichtige Aussagen von allgemeiner Bedeutung gründen.

4. Sinn der Fragebogenaktion im COMCULT-Netzwerk

Angesichts der gerade genannten Schwächen der Fragebogenaktion kann man fragen: Hatte sie denn überhaupt einen Sinn?

Ja, es gibt auch ein paar brauchbare Erkenntnisse, die den Aufwand gerechtfertigt erscheinen lassen:

1. Insgesamt - über alle 365 Pre- und Posttests hinweg – bieten die Ergebnisse eine durchaus brauchbare Beschreibung des Istzustandes an einzelnen Schulen in der EU. Der Wert dieser Feldstudie liegt auch darin, dass sie sich beschränkt auf Inhalte und Werte, die in Schulen gelehrt werden oder gelehrt werden sollten. Gewiss geht die EU mit speziellen Erhebungen in der Reihe EUROBAROMETER oder das infas-Institut, Bonn, mit seiner „Europäischen Wertestudie 1999“ mit 40- bis 50-seitigen Fragebögen weit über den vom

COMCULT-Netzwerk gesteckten Rahmen hinaus. Aber viele Fragen der EU- und infas-Studien zielen auf Dinge, die für den konkreten Schulunterricht über Werte irrelevant sind.

2. Natürlich sind die Fragebogenergebnisse besonders interessant für alle Lehrkräfte, die unmittelbar am Unterricht des COMCULT-Netzwerks beteiligt waren oder in Zukunft entsprechend weiterarbeiten wollen. Sie können vielleicht auch manche offene Frage beantworten, warum in einigen Fällen ihre SchülerInnen anders reagiert haben als ihre Altersgenossen in anderen Ländern.

3. Darüber hinaus fördert der Test vor allem die Einsicht, dass die komplexen Wertefragen in der EU nur noch von einem Lehrerteam umfassend zu unterrichten sind, da ethische, wirtschaftliche und politische Probleme hier eng verzahnt sind.

4. Auch die Notwendigkeit eines Denkens über nationale Grenzen hinweg legen die Fragebogenergebnisse nahe. Das wird nur durch eine massive Unterstützung von Sprachlehrern zu realisieren sein.

5. Alles in allem werfen die Testergebnisse aber auch bildungspolitische Fragen von großer Tragweite auf: Wie können die Schulen neue komplexe Aufgaben lösen in Hinblick auf eine grenzüberschreitende europäische Identität, wenn die Standards für

- die Unterrichtsorganisation
- die Lehrpläne und
- z.T. auch für die Lehrmaterialien

allein von nationalen Regierungen gesetzt werden? Bildungswirklichkeit und Bildungsnotwendigkeit klaffen hier weit auseinander.