Bundesministerium
für Bildung
und Forschung

dipf Deutsches Institut für Internationale Pädagogische Forschung
German Institute for International Educational Research
Member of the Leibniz Association

Conference Volume

Knowledge for action – Research Strategies for an Evidence-Based Education Policy

Symposium during Germany's EU Presidency

28 – 30 March 2007 in Frankfurt/Main

Education and Culture

EU 2007.DE

Deutsches Institut für Internationale Pädagogische Forschung
German Institute for International Educational Research
Schloßstraße 29
60486 Frankfurt am Main
Phone +49 (0) 69.24708 – 0
Fax +49 (0) 69.24708 – 444
E-Mail: dipf@dipf.de
http://www.dipf.de

Inhalt

Preface

The symposium *"Knowledge for Action – Research Strategies for an Evidence-Based Education Policy"* aimed at dealing "with the essential basics of an evidence-based education policy from a research perspective". This is posing a challenge, and successfully dealing with it needs participation of both stakeholders from policy and administration on the one side, and members of the research community on the other side. I do believe that this conference succeeded in bringing together high-level expertise from different countries and from different domains and in setting a starting point for further work.

My gratitude goes to the German federal government and the ministers of the Länder for addressing this important issue in the context of Germany's EU Presidency. The political visibility given to this topic is a first important step towards making an evidence-based policy happen in the future. My thanks also go to the European Commission for setting the stage for a continuous discussion at a European level.

The conference addressed three major areas of work: creation of knowledge, dissemination of knowledge, and application of knowledge. The contributions by the different stakeholders from politics and science validated the importance of these three areas.

Although of equal relevance, there is also a clear hierarchy among them: responsible dissemination and application *require* a scientifically sound basis of knowledge. At the level of indicators of education systems substantial progress has been made over the last years. Nevertheless there are gaps both in production and in dissemination and application. The situation is different concerning our knowledge about what works to overcome observed weaknesses. Evidence on successful interventions is often as weak as our ability to translate existing knowledge into large-scale and politically relevant contexts. The conference pointed to major deficits, but also laid the ground for further work.

At the level of dissemination of knowledge it seems particularly important to me to highlight the efforts made with regard to brokerage agencies. Such agencies seem to be an excellent instrument in dealing with communication problems inherent to any discussion between different communities, as in this case between politics and research. The support by the European Commission and the OECD to establish brokerage agencies in the specific domain of evidence-based education policy is an encouraging indicator for a sustainable process.

Discussions around application of knowledge benefited most from the European dimension of this conference. The different presentations coming from various contexts of several European countries have shown the variety of options decision makers have to engage in considering an evidence-based policy. They have also shown how important further collaboration at an international level will be.

This conference was one step towards a sustainable, internationally collaborative process and I would like to use this opportunity to thank everyone who participated in making this conference a success. I am convinced that the written contributions of the speakers[1] assembled in this publication will be an excellent basis for joint research efforts within and between countries.

Prof. Dr. Eckhard Klieme
Director of the German Institute for International Educational Research

---

[1] For further information on speakers, please visit our website: http://interkoop.dipf.de/index.php?option=com_content&task=view&id=48&Itemid=64

Wissen für Handeln: Die politische Herausforderung

Eröffnungsrede des Parlamentarischen Staatssekretärs bei der Bundesministerin für Bildung und Forschung, Andreas Storm, MdB

Sehr geehrte Damen und Herren,

Ich begrüße Sie im Namen des Bundesministeriums für Bildung und Forschung sehr herzlich zu dieser europäischen, ja internationalen Konferenz zum Thema „Wissen für Handeln–Forschungsstrategien für eine evidenzbasierte Bildungspolitik".

Das BMBF richtet diese Konferenz in Kooperation mit der Europäischen Kommission, namentlich der Generaldirektion „Bildung und Kultur" aus. Ich begrüße sehr herzlich die Generaldirektorin für diesen Bereich, Frau Quintin, sowie den stellvertretenden Generaldirektor, Herrn White. Ihnen sei stellvertretend auch dafür gedankt, dass die Kommission diese Konferenz auch durch finanzielle Unterstützung mit ermöglicht hat. Ich darf ebenfalls sehr herzlich begrüßen Frau Olafsdottir als Vertreterin des Europarates und Leiterin des Schuldepartment und der außerschulischen Bildung der Generaldirektion des Europarates.

Ich danke auch dem Deutschen Institut für Internationale Pädagogische Forschung, seinem wissenschaftlichen Direktor, Herrn Prof. Klieme, seinen Kollegen und Kolleginnen sowie seinen Mitarbeiterinnen und Mitarbeitern für das überaus große Engagement bei der Vorbereitung und Durchführung dieser Tagung.

Meine Damen und Herren,

die turnusmäßige Organisation der Ratspräsidentschaft in der Europäischen Union bringt es mit sich, dass eine Präsidentschaft für ein Land ein relativ „seltenes" Ereignis ist. Dies bedeutet aber zugleich auch, dass die Platzierung eines Themas auf der Agenda der Ratspräsidentschaft eine starke politische Botschaft darstellt – und zwar sowohl nach „innen" auf nationaler Ebene, als auch in seiner Wirkung auf die europäische und internationale Szene.

Mit der heutigen Konferenz nehmen wir die Nahtstelle zwischen Forschung auf der einen und Bildungs- und Forschungspolitik auf der anderen Seite in den Blick. Dass wir für diese Konferenz gezielt das Thema „Wissen für Handeln" ausgewählt haben, mag Ihnen verdeutlichen, wie wichtig aus deutscher Sicht die wechselseitige Beziehung der beiden Bereiche ist.

Bildungspolitik kann die an sie gestellten Erwartungen nur erfüllen, wenn sie sich dabei auf gesicherte Erkenntnisse der Wissenschaft stützen kann. Sie stellt daher hohe Erwartungen an die empirische Bildungsforschung und muss die Rahmenbedingungen für ihre Leistungsfähigkeit schaffen. Noch fehlen aber in vielen Bereichen tragfähige Ergebnisse der empirischen Bildungsforschung. Dies gilt z.B. für die Bereiche der Kompetenzdiagnostik, der Lernprozessgestaltung und Sprachstandsfeststellung, der Schulentwicklungsforschung, der Professionalisierung des pädagogischen Personals, der Systemsteuerung, der Bildungsökonomie und der Indikatorenforschung.

Wir stehen daher in Deutschland gewissermaßen erst am Beginn einer empirischen Wende in Politik und Wissenschaft. Dabei können und müssen wir von anderen Staaten lernen, die mit der Reform ihres Bildungssystems bereits vor längerer Zeit begonnen haben und die Erfolge dieses Vorgehens bei PISA bestätigt bekommen. Aus diesem Grund ist für uns diese Konferenz mit Ihnen von ganz besonderer Bedeutung. So erhoffen wir von Ihnen am Ende der Konferenz Hinweise darauf, was Politik und Forschung leisten sollten, um den Antworten auf viele Fragen ein Stück näher zu kommen: In diesem Sinne danke ich Ihnen,

dass Sie gekommen sind, um diesen Dialog mit uns, aber auch in internationaler Zusammensetzung zu führen.

Für uns wichtig zu wissen ist z.B., inwieweit es in Ihren Ländern gelingt, Evidenz in den Bereichen von Politik und Wissenschaft herzustellen. Welche Instrumente sind dafür vonnöten? Aber auch: Wo sind die Grenzen dessen, was die Wissenschaft an Evidenz für die Politik überhaupt herstellen kann? Welche Bedeutung haben Indikatoren- und Benchmarksysteme, und in welche Richtung gilt es, sie weiterzuentwickeln? Welche Voraussetzungen müssen auf Seiten der Politik wie der Forschung dafür geschaffen werden? Was leisten Interventionsstudien in diesem Zusammenhang? Welche Art von Wissensmanagementsystemen und Vermittlungsagenturen sind vonnöten, um den Vermittlungsprozess zwischen Politik und Wissenschaft anzubahnen und zu unterstützen?

Selbstverständlich geht das grundlegende Thema evidenzbasierten Entscheidens und Handelns über die Beziehung zwischen Forschung und Politik weit hinaus. Akteure aus der Praxis, Lehrer und Schulleiter, aber auch Eltern sind genauso Adressaten wissenschaftlicher Erkenntnis wie die politischen Amtsträger. Umgekehrt braucht Politik gerade die Anregung dieser Akteure. Die jeweiligen Wege von wissenschaftlicher Erkenntnis zu praktischem Handeln und umgekehrt sind für unterschiedliche Adressaten so unterschiedlich, dass ich in meinen weiteren Ausführungen zur Frage nach der spezifisch „politischen Herausforderung“ vor allem folgende Schwerpunkte setzen möchte:

1. Welche Ausgangslage führt dazu, dass für uns die Nahtstelle zwischen Forschung und Politik überhaupt so relevant geworden ist?
2. Welches sind die legitimen und notwendigen bildungspolitischen Erwartungen der Politik an Forschung, speziell an die Empirische Bildungsforschung?
3. Welches sind aber auch die notwendigen Verpflichtungen der Politik gegenüber Bildungsforschung?

Zu 1: Zur Ausgangslage

Zur Skizzierung der Ausgangslage: Wir werden schnell Konsens dahingehend erzielen, dass die unbestreitbaren Erfolge einer wissensbasierten und output-orientierten Steuerungsphilosophie zu einem Nachdenken darüber geführt haben, ob wir nicht einen Paradigmenwechsel hin zu einer „empirischen Wende“ in Politik und Wissenschaft brauchen. Kernelemente dieses Paradigmenwechsels finden wir schnell:

Im Wesentlichen in einem sinnvoll aufeinander abgestimmten System von regelmäßigen Schulevaluationen, von nationalen und internationalen Leistungsuntersuchungen, einer unabhängigen und wissenschaftlichen Bildungsberichterstattung sowie einer leistungsfähigen Bildungsstatistik, die wiederum eine leistungsfähige empirische Bildungsforschung als Leitdisziplin voraussetzt. Während erfolgreiche PISA-Teilnehmerländer diese „empirische Wende“ in beiden Segmenten – in der Politik wie in der Forschung – spätestens in den 1980er Jahren erfolgreich und vor allem zielgerichtet begonnen haben, setzte diese Entwicklung in Deutschland mit Entschiedenheit erst nach der Veröffentlichung der Ergebnisse von PISA 2000 ein.

Derzeit sind wir auf dem Weg, einem guten Weg, möchte ich betonen. Für Deutschland sind insbesondere die skandinavischen Staaten, die Niederlande, aber auch Kanada Beispiele dafür, wie dieser Weg erfolgreich beschritten werden kann. Nicht nur, dass Forschung z.B. in Form von großen Längsschnittstudien und Assessments die Grundlagen und die Instrumentarien für neue bildungspolitische Ansätze liefert. Auch umgekehrt führt gerade der praktische Einsatz solcher Forschungsinstrumente wiederum dazu, dass neue Impulse für eine Weiterentwicklung der Forschungsinstrumente und der möglichen Fragestellungen entstehen.

Der Nutzen einer solchen wissensbasierten und zugleich output-orientierten Steuerung liegt darin, dass im Bereich der Bildungspolitik Mechanismen installiert werden, die

- automatisch auf Erfolge und Fehlentwicklungen gleichermaßen aufmerksam machen,
- somit den „Zwang zum Lernen“ im System selbst verankern und
- letzteren insbesondere von politischer Opportunität lösen.

Noch fehlt die Forschung, die zielgenau den Zusammenhang zwischen den Leistungsergebnissen einerseits und wissensbasierter sowie output-orientierter Steuerung andererseits wirklich nachweist. Wir sind an Forschung in diesem Feld hoch interessiert und werden dieses Thema auch weiterhin intensiv verfolgen.

Eine solchermaßen empirisch fundierte Bildungspolitik ist auf ein zukunftsfähiges Bildungsmonitoring angewiesen, das hinsichtlich seiner Leistungsfähigkeit international anschlussfähig sein muss. Es muss dabei zugleich der Begrenztheit finanzieller und wissenschaftlicher Ressourcen Rechnung tragen und in kluger Weise Zuständigkeitsfragen auch unter den Notwendigkeiten gesamtstaatlicher Steuerung behandeln.

Zu 2: Erwartungen der Politik an Bildungsforschung

Damit komme ich zu meinem zweiten Punkt: Die politischen Erwartungen an Bildungsforschung richten sich darauf, Antworten auf all die Fragen zu erarbeiten, die im Zuge einer wissensbasierten Systemsteuerung entstehen. Beispielhaft seien genannt:

- Erkenntnisse über „Bildung im Lebenslauf“,
- empirisch prüfbare Bedingungen für das Gelingen von guten Lehr-Lernprozessen,
- Gelingensbedingungen von Schulentwicklungsprozessen,
- Grundlagen und Verfahren der Kompetenzdiagnostik,
- Indikatorenforschung zur Gewinnung von aussagefähigen Daten für eine indikatorengestützte Bildungsberichterstattung
- und vieles mehr.

Auch von einer Forschung, die praktische Probleme aufgreift, muss eine theoretische Fundierung erwartet werden. Sie muss aufzeigen, welche theoretisch relevanten Phänomene sie erklärt und welche Theorien durch ihre Ergebnisse erweitert werden können und müssen.

Eine so verstandene „nutzenorientierte Forschung“ intendiert dezidiert praktische Effekte, und sie zieht diese auch zur Begründung und Beschreibung neuer Fragestellungen heran.

Eine weitere Daueraufgabe ist meines Erachtens darüber hinaus, Forschungsergebnisse so aufzubereiten, dass Bildungspolitik ihr vorhandenes Wissen sinnvoll damit verknüpfen kann. Ergebnisse der Forschung müssen so aufbereitet werden, dass sie in gesellschaftliche Diskussionen und politische Entscheidungen hinein kommuniziert und dort auch angewendet werden können. Dies zu gewährleisten ist ebenfalls eine zentrale Maßnahme von Qualitätssicherung – aus Sicht der Politik wie der Wissenschaft.

Zu 3. Verpflichtungen der Politik gegenüber Forschung

Damit Wissenschaft und Forschung dieses und vieles mehr leisten können, bedarf es zunächst klarer politischer Entscheidungen über Maßnahmen zur strukturellen Stärkung der Forschung, speziell der Empirischen Bildungsforschung. Das Bundesministerium für Bildung und Forschung hat hier klare Signale gesetzt.

Im Sommer dieses Jahres werden wir ein „Rahmenprogramm zur strukturellen Stärkung der empirischen Bildungsforschung“ auflegen, das einerseits gezielt Maßnahmen zur strukturellen Stärkung der empirischen Bildungsforschung enthält und andererseits thematisch orientierte Förderschwerpunkte vorsieht. Das Rahmenprogramm wird dabei unterschiedliche Handlungsoptionen strategisch verbinden, um die Empirische Bildungsforschung in Deutschland strukturell und damit langfristig zu stärken.

Wesentliche Maßnahmen zur strukturellen Stärkung betreffen zum einen Qualitätsentwicklung und -sicherung in der vom BMBF – bzw. gemeinsam vom BMBF und den Ländern – geförderte Empirische Bildungsforschung. Ein wichtiges Element für die Förderung und Herausbildung von Exzellenz stellen Wettbewerbselemente dar. Sie werden deshalb in der künftigen Förderpraxis erheblich starker – z.B. in Form von Ausschreibungen – zum Tragen kommen. Auch die Trennung von Begutachtung und Förderentscheidung wird ein wesentliches Element in diesem Zusammenhang sein.

Gezielte Maßnahmen zur Nachwuchsförderung sollen – in Abstimmung mit maßgeblichen Forschungsfördereinrichtungen wie der Deutschen Forschungsgemeinschaft – sicherstellen, dass auch innerhalb der BMBF-geförderten Projekte und Maßnahmen aktive Nachwuchsförderung betrieben wird. Darüber hinaus ist aber auch eine explizite Nachwuchsförderung in Form von ausgeschriebenen Promotionsstipendien zu definierten Themenbereichen vorgesehen.

Ein besonderes Augenmerk werden wir der Verbesserung der informationellen Infrastruktur widmen. Gute Förderbedingungen beziehen sich auch auf den Zugang und die Verfügbarkeit von Daten. Das Rahmenprogramm wird deshalb Vorschläge enthalten zur

- Sicherung des Zugangs zu Daten aus öffentlich geförderten Forschungsvorhaben über Forschungsdatenzentren,
- zu Regelungen der verpflichtenden Archivierung und Weitergabe (anonymisierter) Mikrodaten zu wissenschaftlichen Zwecken bei BMBF-geförderten Projekten sowie,
- zur Erschließung bisher nicht zugänglicher Datenquellen für die Bildungsforschung.

Weiter ist für uns die Entwicklung einer Kommunikations- und Veröffentlichungsstrategie für den Bereich Bildungsforschung ein besonderes Anliegen. Ergebnisse der Bildungsforschung, die aus vom BMBF initiierten und finanziell unterstützten Forschungsvorhaben stammen, sollen künftig noch stärker als bisher sowohl für ein Fachpublikum als auch für ein bildungspolitisch interessiertes Publikum zugänglich gemacht werden. Dies soll keineswegs – dies zu betonen ist mir ein ausdrückliches Anliegen – die Publikation von wissenschaftlichen Ergebnissen in internationalen Zeitschriften behindern oder damit konkurrieren.

Neben den Maßnahmen zur strukturellen Stärkung wird das BMBF seine Förderpraxis an thematischen Förderschwerpunkten ausrichten. Hierfür werden wir unsere Projekt- und Programmförderung zu thematischen Schwerpunkten bündeln, bei denen aus bildungspolitischen Gründen besonderer Erkenntnisbedarf besteht. Wir werden damit auch die Voraussetzungen für deutsche Bildungsforscher verbessern, Debatten über prioritäre Themen der internationalen Forschungsagenda sowie die Ausgestaltung internationaler Programme, z.B. im Bereich der OECD und der EU, mehr noch als bisher mitzugestalten.

Zu den bereits angelaufenen größeren Vorhaben gehört unter anderem die Mitwirkung bei der Etablierung und ggf. Kofinanzierung eines von der Wissenschaft getragenen nationalen Bildungspanels. Es soll einerseits mittel- und langfristig eine aussagekräftige Datenbasis für eine an „Bildung im Lebenslauf“ orientierte Bildungsberichterstattung darstellen.
Andererseits wird es einen entscheidenden Beitrag zur Verbesserung der strukturellen Rahmenbedingungen für die empirische Bildungsforschung in Deutschland leisten.

Fragestellungen, die mit einem Panel bearbeitet werden können und sollen, sind unter anderem

- die Analyse von Kompetenzentwicklung im Lebenslauf innerhalb und außerhalb von Bildungs- und Ausbildungsinstitutionen sowie mit Bezug auf familiäre, soziale und gesellschaftliche Kontexte,
- die Analyse von Bildungsentscheidungen und Bildungsprozessen bei kritischen Übergängen,
- die Relevanz spezifischer Kompetenzen für Ausbildungs- und Berufserfolg sowie
- die Leistungen von Bildungs- und Ausbildungseinrichtungen, gemessen am Erfolg ihrer Absolventen.

Ein zweiter bereits laufender Forschungsschwerpunkt betrifft Fragen der Kompetenzdiagnostik: Eng verknüpft mit der Konzeption eines Bildungspanels, das „Kompetenzentwicklung im Lebenslauf" zum Thema hat, ist die Frage, welche Kompetenzen in welchem Alter wie gemessen werden sollen und auch tatsächlich valide gemessen werden können. Das BMBF hat deshalb im Rahmen seiner Forschungsförderung die Voraussetzungen dafür geschaffen, dass sich hier Schwerpunkte etablieren konnten: Zum einen in Form eines DFG-Schwerpunktprogramms „Kompetenzdiagnostik", in dem grundlegende Fragen zur Kompetenzdiagnostik behandelt werden sollen, zum anderen in Form einer begleitenden Förderinitiative des BMBF zu Fragen der technologiebasierten Kompetenzdiagnostik.

Künftige thematische Förderschwerpunkte werden so beschaffen sein, dass die mit ihnen verbundenen Forschungsfragestellungen bildungsbereichsübergreifend aufgegriffen werden können. Zudem sollen sie an aktuelle und künftige Problembereiche des Bildungssystems anknüpfen und mit dazu beitragen, die Wissensbasis für eine evidenzbasierte Systemsteuerung zu verbreitern. Soweit es sich um Forschungsvorhaben in Bildungsbereichen handelt, für die die Zuständigkeit bei den Ländern liegt, – Herr Meyer-Hesemann weiß dies – wird dies selbstverständlich in Abstimmung mit den Ländern erfolgen.

Schließlich: Zentral für die internationale Anschlussfähigkeit der deutschen (empirischen) Bildungsforschung ist ihre internationale Sichtbarkeit und Vernetzung. Das Rahmenprogramm wird deshalb auch Vorschläge enthalten, mit welchen Instrumenten das BMBF eine weitergehende Internationalisierung fördern will. Zu nennen sind hier insbesondere Stipendien für Forschungsaufenthalte von Nachwuchswissenschaftlern im Ausland und eine zusätzliche Unterstützung bei der Publikation von Forschungsergebnissen in internationalen Zeitschriften.

So fördert das BMBF seit kurzem beim Deutschen Institut für Internationale Pädagogische Forschung ein Büro für Internationale Kooperation in der Bildungsforschung. Es soll Unterstützung leisten bei der Entwicklung strategischer Überlegungen, die dazu dienen, vor allem im Bereich evidenzbasierter Politik die Vernetzung auf europäischer Ebene und die Anbahnung internationaler Kooperationsvorhaben und tragfähiger Forschungsstrategien voranzutreiben. Diese Konferenz ist insofern aus deutscher Sicht ein erster Auftakt für diesen notwendig weiterzuführenden Dialog. Mit dieser Zusage möchte ich schließen und Ihren Beratungen gutes Gelingen wünschen und hoffentlich auch eine gute Resonanz.

Wissen für Handeln – Forschungsstrategien für eine evidenzbasierte Bildungspolitik

Rede des Staatssekretärs für Bildung und Frauen Schleswig-Holstein,
Wolfgang Meyer-Hesemann

Herr Professor Klieme,
Herr Kollege Storm,
meine sehr geehrten Damen und Herren,

Bildung nimmt eine Schlüsselrolle für die individuelle Entwicklung, für gesellschaftliche Teilhabe sowie für berufliches Fortkommen, aber auch für den wirtschaftlichen Erfolg eines Landes ein. Die globalen Entwicklungen der vergangenen Jahre haben diese grundlegende Bedeutung von Bildung für alle Staaten Europas unter der Perspektive einer Entwicklung hin zu wissensbasierten Gesellschaften noch einmal unterstrichen.

Um so wichtiger ist es, und dafür danke ich den Initiatoren dieser Veranstaltung, sich darüber zu verständigen, wo Möglichkeiten und Grenzen einer durch empirische Daten fundierten Bildungspolitik liegen. Einer Bildungspolitik, bei der Entscheidungen auf der Grundlage von Wissensbeständen und nicht länger nach „gefühlten Einschätzungen" getroffen werden.

Sie, sehr verehrte Damen und Herren, werden bis zum Freitag über die dafür erforderlichen Forschungsstrategien diskutieren. Ich darf Ihnen - auch im Namen des Präsidenten der Kultusministerkonferenz, Senator Prof. Jürgen Zöllner - die Grüße der Bildungs- und Wissenschaftsminister aller 16 Länder der Bundesrepublik Deutschland übermitteln und diesem Kongress viel Erfolg wünschen.

Ich werde Ihre Zeit nicht beanspruchen für eine detaillierte Nabelschau der Probleme des für Außenstehende eher skurrilen deutschen Bildungsföderalismus. Aber ich möchte gerne aus Sicht der Länder auf die grundsätzliche Bedeutung evidenzbasierter Bildungspolitik eingehen. Und ich möchte etwas zu den Erwartungen sagen, die wir in diesem Zusammenhang an die Bildungsforschung haben. Dazu gehört es, auch einige noch zu klärende Fragen zu stellen.

Waren es zu früheren Zeiten die Langzeitfolgen des „Sputnik-Schocks", die ein Umdenken in der Bildungspolitik auslösten, haben in den vergangenen Jahren vor allem die Ergebnisse internationaler Schulleistungsstudien in Deutschland einen neuen „Bildungsschock" ausgelöst. Das unbestreitbar große und – ich glaube, man kann es im Abstand weniger Jahre schon sagen – das historische Verdienst der empirischen Bildungsforschung sehe ich darin, den im selbstzufriedenen Glauben an seine Überlegenheit und Vorbildfunktion wichtige internationale Entwicklungen verschlafenden Riesen mit dem Namen „Deutsches Bildungssystem" wachgerüttelt zu haben.

Es waren die 1997 publizierten unerwartet niedrigen Leistungsergebnisse der TIMSS-Studie und dann 2001 die PISA-Ergebnisse, die unsere Annahme erstmalig nachhaltig erschütterten, dass deutsche Schülerinnen und Schüler im internationalen Vergleich relativ gut abschneiden würden. Deutschland hat seitdem von der, wenn auch verspäteten, Einbindung in den internationalen Kontext empirischer Bildungsforschung und der Beteiligung an internationalen Large Scale Assessments in besonderer Weise profitiert.

Spätestens seit Veröffentlichung von PISA 2000 ist eine heftige Gegenbewegung im Pendel der öffentlichen Einschätzung der Qualität des deutschen Bildungssystems festzustellen. Teilaspekte werden vorschnell fürs Ganze genommen und das Kind droht mit dem Bade ausgeschüttet zu werden. Die einstmals vermeintlich führende Bildungsnation ist

mittlerweile weithin davon überzeugt, im internationalen Vergleich abgeschlagen hinten zu liegen und hat sich an das schlechte Selbstbild mit einem gewissen Masochismus schon fast gewöhnt.

Kennzeichnend für jeden Aufholprozess ist offenbar ein hohes Maß an Aufgeregtheit bei den verspätet Startenden. Sie wollen in wenigen Jahren alles nachholen, was lange versäumt wurde und neigen daher zu Übertreibungen und auch großer Ungeduld. Ich glaube, das ist eine Erklärung dafür, weshalb wir in Deutschland von der selbstbewussten und gleichzeitig konzentrierten Gelassenheit mancher europäischer Nachbarn, im Umgang mit den Ergebnissen internationaler Studien noch ein ganzes Stück entfernt sind.

Dennoch bin ich optimistisch, dass sich das Pendel unserer öffentlichen Selbsteinschätzung langsam wieder auf ein realistisches Maß einrichten wird und dass wir die Ergebnisse kommender internationaler wie nationaler Schulleistungsvergleiche weniger aufgeregt und mit einem vertieften Interesse für den inhaltlichen Gehalt desto gründlicher auswerten werden.

Weil es manchmal in Vergessenheit gerät, sei es hier noch einmal in Erinnerung gebracht: Die deutsche Bildungspolitik hat nicht nur auf die TIMSS-, PISA- und IGLU-Schockwellen umfassend reagiert. Sie hat diese mit ausgelöst. Die Bildungsminister der Länder selbst haben in der Kultusministerkonferenz mit den sogenannten Konstanzer Beschlüssen vom Oktober 1997 die Qualitätssicherung im deutschen Schulwesen im internationalen Vergleich zu ihrem zentralen Thema gemacht hat.

In allen Ländern haben TIMSS, PISA und IGLU inzwischen die viel beschworene „empirische Wende“ ausgelöst, um einen Begriff Jürgen Baumerts zu verwenden. Eine Wende, die zu einer an den Ergebnissen von Bildungsprozessen ausgerichteten Steuerungsphilosophie geführt hat – bei gleichzeitiger Erweiterung der Handlungsspielräume für die einzelne Schule. Es ist heute evident, dass die bis dahin vorrangige Inputsteuerung nicht in dem erhofften Maße zur erwünschten Qualität im Bildungssystem beigetragen hat. Seitdem gilt für alle Länder, dass weder eine ernst zu nehmende Bildungspolitik noch eine erfolgreiche Schulgestaltung ohne empirisch abgesichertes, praxisrelevantes Steuerungswissen vorstellbar sind.

Die Länder haben in den vergangenen 10 Jahren große Anstrengungen unternommen, um die notwendigen Grundlagen für eine evidenzbasierte Bildungspolitik zu schaffen. Die zentralen Elemente eines länderübergreifenden nationalen Bildungsmonitoring wurden von der Ländergemeinschaft trotz mancher Konflikte in einem relativ knappen Zeitraum erfolgreich aufgebaut. Dazu gehören die Festlegung von Bildungsstandards und die Entwicklung von Testverfahren zu deren Überprüfung, nationale wie internationale Leistungsvergleiche und eine systematische, indikatorenbasierte Bildungsberichterstattung.

Seit vielen Jahren werden mit neuer Intensität – und hier zitiere ich gerne Eckhard Klieme – „Fragen der Gesamtproduktivität des Bildungssystems (...) ebenso untersucht wie die Chancen zur persönlichen Entfaltung und Bildung von Kindern und Jugendlichen, die Qualität von Schule und Unterricht, das Zusammenspiel von sozio-biographischen und schulischen Faktoren und nicht zuletzt Fragen der Chancengerechtigkeit im Bildungssystem“ (Eckhard Klieme in: „Impulse für die Bildungsforschung- Stand und Perspektiven, DFG, S. 99).

Die Länder haben darüber hinaus erkannt, dass sie zur Umsetzung dieser umfassenden Aufgaben nicht nur mit den vorhandenen Forschungs-Instituten eng zusammen arbeiten, sondern auch für eine eigene Serviceeinrichtung sorgen müssen. Mit der Gründung des Instituts für Qualitätsentwicklung im Bildungswesen (IQB) als An-Institut an der Humboldt-Universität in Berlin ist dieser Schritt erfolgreich gelungen. Das IQB hat sich mit seinem Leiter, Olaf Köller, und den exzellenten Mitarbeitern innerhalb kurzer Zeit einen hervorragenden Ruf erarbeitet. Es unterstützt neben den Arbeiten zur Entwicklung, Normierung und Überprüfung von Bildungsstandards die Länder in allen Fragen evidenzbasierter Bildungspolitik.

Die Kultusministerkonferenz hat im Juni 2006 außerdem eine Gesamtstrategie zur Beobachtung des Bildungssystems vorgelegt. Sie dient dazu, die vordringlichen Felder der Bildungsforschung zu definieren, dem steigenden Bedarf an abgesichertem Steuerungswissen gerecht zu werden und um die bisher fehlende systematische Verknüpfung der verschiedenen Evaluations- und Steuerungs-Ebenen zu gewährleisten. Dabei geht es nicht nur um die systematische Beschaffung von Informationen zur Qualität des Bildungssystems. Es geht gleichzeitig um die enge Verknüpfung dieser Informationen mit Maßnahmen zur Unterrichts- und Qualitätsentwicklung, die der konkreten Arbeit an jeder einzelnen Schule zugute kommen sollen.

Die Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring umfasst vier konzeptionell miteinander verbundene Bereiche, die ich nur kurz erwähnen möchte:

- Internationale Schulleistungsuntersuchungen
- Die zentrale Überprüfung der nationalen Bildungsstandards in einem Ländervergleich
- Vergleichsarbeiten zur landesweiten Überprüfung der Leistungsfähigkeit einzelner Schulen
- Die gemeinsame Bildungsberichterstattung von Bund und Ländern.

Im Verständnis der Länder sind alle Verfahren der Bildungsforschung kein Selbstzweck. Sie dienen im Kern dazu, „Bildungswirklichkeit zu verstehen und zu verbessern“, wie es Manfred Prenzel kurz und zutreffend auf den Punkt gebracht hat.

Wenn wir diese Vorgabe ernst nehmen, müssen wir 10 Jahre nach TIMSS auch die bisherige Entwicklung der empirischen Bildungsforschung in Deutschland kritisch bilanzieren. Aus meiner Sicht werden dabei neben unbestreitbaren Erfolgen auch Defizite deutlich.

Sicherlich sind auch weiterhin Erhebungen abgesicherter und steuerungsrelevanter Daten zur Kompetenzentwicklung und zu den Wirkungen bildungspolitischer Maßnahmen in zentralen Bildungsbereichen erforderlich. Daneben wächst in den Ländern aber vor allem das Interesse an Studien, die praktikable Wege zur Lösung der beschriebenen Probleme aufzeigen. Insbesondere geht es um abgesichertes und anwendbares Steuerungswissen zur zentralen Herausforderung, wie eine erhöhte Bildungsqualität bei gleichzeitiger Verbesserung der Bildungschancen erreicht werden kann.

Unter dieser Prämisse müssen die bisherigen Forschungsansätze ergänzt werden um Forschungsvorhaben, die Erklärungswissen hinsichtlich kausaler Einflussgrößen zentraler bildungspolitischer Probleme bieten und um Interventionsstudien, die Veränderungswissen bereitstellen, inwieweit einzelne Maßnahmen geeignet sind, um bestimmte Probleme zu lösen.

Das besondere Forschungsinteresse der Länder gründet sich im Übrigen auf den zentralen, von der Bildungsforschung diskutierten Desiderata der letzten Jahre. Es lässt sich in vier Themenfeldern zusammenfassen, die sich im Schwerpunkt auf das Geschehen im Unterricht beziehen.

1. Die Weiterentwicklung und Analyse von Tests und Verfahren der Evaluation

Die Bereitstellung von kostengünstigen, einfach verfügbaren und effektiven Instrumenten für die kompetenzorientierte Individualdiagnose, die Lehrkräften Handlungswissen für das Unterrichtsgeschehen bereit stellen, steht noch am Anfang und sollte ausgebaut werden.

2. Wirksame Impulse zur Unterrichtsentwicklung und Schulentwicklung

Sowohl die Large-scale Assessments der letzten Jahre als auch zahlreiche Unterrichtsstudien haben deutlich gemacht, dass erfolgreiche Wissenserwerbsprozesse in einem handlungs-orientierten, kognitiv aktivierenden Unterricht stattfinden. Als Folge der Einführung von

Bildungsstandards und der damit verbundenen Fokussierung auf einen kompetenzorientierten Unterricht ist es daher wünschenswert, die Anstrengungen im Bereich der Unterrichtsforschung zu intensivieren.

3. Erfolgreiche Wege zur individuellen Förderung von Schülerinnen und Schülern

Die großen sozialen und ethnischen Disparitäten in der Bildungsbeteiligung und in den Schulleistungen haben sich als zentrale Herausforderungen für die deutsche Bildungspolitik herausgestellt. Individuelle Förderung muss alle Schülerinnen und Schüler erreichen, insbesondere aber auf die des unteren und des oberen Leistungsbereichs ein besonderes Augenmerk haben.

4. Zielgerichtete Professionalisierung der Lehrkräfte

Erfolgreiche Lehr-/Lernprozesse hängen notwendigerweise in einem hohen Maße von der Professionalität der unterrichtenden Lehrkräfte ab. Die dafür erforderliche Lehrerexpertise bezieht sich auf fachliche, fachdidaktische und pädagogische Fragen. Vorliegende Forschungsergebnisse verweisen darauf, dass die Kombination aller drei Bereiche zu wünschenswerten Lernergebnissen auf Seiten der Schülerinnen und Schüler führt.

Dieses Forschungsinteresse werden die Länder gemeinsam mit dem Bund abstimmen. Dafür bietet die im Zuge der Föderalismusreform neu eingerichtete Gemeinschaftsaufgabe von Bund und Länder zur „Feststellung der Leistungsfähigkeit im Bildungswesen im internationalen Vergleich“ eine gute Grundlage.

Noch nie zuvor waren die Erwartungen der Bildungspolitik an die Bildungsforschung in Deutschland so hoch wie heute und noch nie stand Bildungspolitik so sehr unter einem Rechtfertigungsdruck gegenüber den Ergebnissen empirischer Bildungsforschung. Lassen Sie mich daher zum Schluss auf das komplizierte, aber wie ich finde, spannende Verhältnis von Bildungspolitik und Bildungsforschung eingehen, über das Sie auch im Rahmen Ihrer Tagung diskutieren werden.

Herr Kollege Storm vom BMBF hat Ihnen erläutert, auf welchen Wegen und auf welche Weise das Bundesbildungsministerium die empirische Bildungsforschung unterstützen wird. Dazu gehören: Qualitätsentwicklung und -sicherung in der Bildungsforschung, eine verbesserte informationelle Infrastruktur und Nachwuchsförderung, eine stärkere Internationalisierung und eine bessere Kommunikations- und Veröffentlichungsstrategie. Diese sinnvollen und notwendigen Ansätze werden von den Ländern befürwortet.

Alle Beteiligten sollten sich aber vergegenwärtigen, dass die von Bund und Ländern geplanten Maßnahmen zur Förderung und Unterstützung der Bildungsforschung nur einen Impuls neben anderen setzen können. Es darf nicht der Eindruck entstehen, als wollten Bund und Länder die empirische Bildungsforschung auf die Funktion reduzieren, Auftragsforschung zur Generierung von Steuerungswissen für die Bildungspolitik zu betreiben. Bildungsforschung ist mehr als Bestellforschung und braucht Freiräume für Grundlagenforschung und eigene Schwerpunkte. Z.B. auch zur Beantwortung der brisanten Frage, wie Bildungspolitik mit den Ergebnissen empirischer Bildungsforschung umgeht oder auch zur Aufdeckung politikinduzierter Probleme im Bildungsbereich (man könnte zugespitzt sagen: im Sinne einer „Sience of bad practice“) – worauf Heinz-Elmar Tenorth jüngst hingewiesen hat.

Das Motto der Tagung – „Wissen für Handeln“ – darf nicht die falsche Erwartung wecken, wissenschaftlich abgesichertes Steuerungswissen ließe sich unmittelbar kraft Rationalität der Erkenntnis in den Beschluss von bildungspolitischen Maßnahmen umsetzen. Diese Erwartung ist naiv, denn sie verkennt die vollkommen unterschiedlichen handlungsbestimmenden Koordinaten politischen Handelns und wissenschaftlichen Arbeitens: Wissenschafter stellen komplexe Fragen, für deren Beantwortung sie möglichst

viel Zeit und Mittel benötigen; Politiker suchen klare und möglichst einfache und umsetzbare Antworten zur Lösung der aufgezeigten Probleme wie zur Durchsetzung ihrer politischen Zielvorstellungen.

Produktiv ist das Verhältnis immer dann, wenn Wissenschaftler in der Lage sind, der Bildungspraxis abgesicherte Forschungsergebnisse zu relevanten Fragen vorzulegen, die auch tatsächlich genutzt werden können, und wenn Politiker die durch Forschungsergebnisse abgebildete Realität ungeschminkt zur Kenntnis nehmen, um daraus die notwendigen Schlussfolgerungen zu ziehen. Schwierig wird es, wenn Politiker sich als Wissenschaftler verhalten und Wissenschaftler als Politiker.

Empirische Bildungsforschung erfüllt nur dann ihren umfassenden Auftrag, wenn die Wissenschaftlerinnen und Wissenschaftler ihre Unabhängigkeit immer wieder behaupten, unbequem bleiben und sich von den Ansprüchen politischer Instrumentalisierung frei machen. Ich weiß wovon ich rede: Auch wenn es Bildungspolitikern manchmal schwer fällt, brauchen wir den unbestechlichen und klaren Spiegel empirischer Evidenz. Bildungspolitiker und Wissenschafter müssen sich daher in einen kritischen und offenen Dialog begeben, der nicht ohne Reibungen und Konflikte verlaufen kann.

Ich begrüße die Initiative zur Ausrichtung dieser Tagung nicht nur aus diesem Grund, sondern auch, weil die in den verschiedenen Workshops angebotenen Themen eine gute Möglichkeit bieten, um offene Fragen auf dem Weg zu einer evidenzbasierten Bildungspolitik zu diskutieren. Aus meinen Erfahrungen der vergangenen Jahre will ich zum Schluss vier mir besonderes wichtige offene Fragen nennen:

1. Wo liegen die Möglichkeiten und Grenzen von Benchmarks und Vergleichen von Staaten, Ländern und einzelner Schulen? Wie stärken wir die diagnostische wie die pragmatische Funktion dieser Verfahren und vermindern das Interesse an vordergründigen Hitlisten? Wie verhindern wir ein teaching to the test - und zwar auf allen Ebenen - vom Bildungspolitiker bis zur einzelnen Lehrkraft?
2. Wie können wir eine vertiefte und verbreiterte Diskussion der Ergebnisse empirischer Bildungsforschung in der Öffentlichkeit erreichen, die dem Gehalt der Studien angemessen ist? Wie können wir die Ergebnisse besser übersetzen, so dass alle Beteiligte davon in einem stärkeren Maße als bisher profitieren?
3. Wie verbessern wir die Verbindungen zwischen Bildungsforschung, Bildungspolitik und Bildungspraxis, so dass praxisrelevante Ergebnisse vorgelegt und genutzt werden? Wie können wir Lehrkräfte und Bildungsforscher zum gegenseitigen Nutzen besser zusammenbringen? Und schließlich:
4. Wie stimmen wir die verschiedenen Studien auf nationaler wie internationaler Ebene besser ab und kommen zu einer vernünftigen und leistbaren Gesamtkonzeption eines internationalen Bildungsmonitoring? Wie sichern wir innerhalb und zwischen den Studien eine Kontinuität entsprechender Indikatoren, um Längsschnitte und Vergleiche ableiten zu können?

Ich hoffe, mit diesen Fragen ein wenig auch Ihren Erkenntnisnerv getroffen zu haben und wünsche Ihnen einen reichen wissenschaftlichen, bildungspolitischen und auch ganz praktischen Ertrag von dieser Tagung.

Ich danke für Ihre Aufmerksamkeit

The Strategy of the European Union

Odile Quintin, Director General of DG Education and Culture, European Commission

Dear colleagues, ladies and gentlemen,

I want to address my first words today to our hosts, the German Presidency.

First, my thanks for organising this conference. It answers the calls from European Education Ministers to take forward the issue of evidence-based policymaking. This is a somewhat more complicated issue than appears at first sight. The conference will be very helpful in clearing the path ahead.

Secondly, my thanks for Sunday's declaration in Berlin, which brought home the tangible benefits of 50 years of peaceful development in Europe. Berlin itself is a human story of separation and unity. One which tells us that Europe has the determination and the capacity to overcome divisions.

Finally, my thanks for the Presidency's support for some highly strategic initiatives. Chief among these is our proposal for a European Institute of Technology, strengthening the relationship between education, research and innovation. We are in constant contact, to get the EIT off the ground as quickly as possible. We are all committed to giving a clear signal that the innovation challenge must be tackled. The EIT should be a milestone in achieving this goal.

The German Presidency also sees the launch of our new programmes in lifelong learning, youth and culture. A full Presidency programme, in other words, and one which you are dedicated to achieving.

Why the need for evidence-based policy?

Now I want to turn towards the theme of this conference. Why is an evidence base so important for education policy? And how can the Commission support Member States?

The answer is clear. Education and training are at the core of Member State and European strategies for meeting the challenges of the 21st century. The Berlin declaration makes it plain: Europe's wealth is the knowledge and ability of its people; that is the key to growth, employment and social cohesion.

Education and training are indispensable building-blocks of the knowledge economy. They are intimately linked with the key issues across Europe: maintaining cohesion, active ageing, integrating migrants, reducing wealth disparities.

Europe's future therefore depends to a large degree on making the right decisions in education and training. This is no over-statement: the stakes are high.

Such high stakes means that decision-makers and practitioners are under intense scrutiny, from the public, the media and politicians. Policy decisions are taken in a highly-charged atmosphere. Practitioners have to account for themselves, to their authorities, to parents, employers and their communities.

But in order to make the right decisions, we must know what works. There is consensus that we need to bring knowledge – research, data – and action – teaching and learning practice – closer together. Working in harness, knowledge and action can power the debate that shapes policy.

Education and training, however, do not live in a vacuum. They are part of the cultural identities of countries and regions. They interact with a web of other policies.

In these circumstances, there are no simple prescriptions about what makes up relevant “evidence”. We need to acknowledge from the start that individual pieces of research are context-specific. They are a source of insight rather than the whole picture. Comparing different systems can be more difficult than it seems. The process is dynamic and short on fixed points.

But still, I see enormous value in tapping into the experiences that exist at European level. We can use the growing body of evidence to learn from each other, to learn what works and when.

Je me permets maintenant de passer en français pour en dire plus.

Le programme de travail "éducation et formation 2010"

Bien sûr, les débats d'aujourd'hui doivent être replacés dans le contexte de la stratégie de Lisbonne et du Programme "Education et Formation 2010". C'est notre fil conducteur. Dans ce cadre, les États Membres poursuivent un vaste processus de modernisation de leurs systèmes éducatifs. Nous avançons ensemble dans le respect des compétences de chacun. En fait, nos objectifs très ambitieux dans un domaine où les compétences sont partagées ont nécessité la mise en place d'une méthode de coopération originale. Les Etats membres sont incités a évaluer leurs actions et échanger leurs bonnes pratiques en vue d'améliorer la qualité, l'efficacité et l'équité de leurs systèmes.

Cette méthode implique aussi des prises de décision qui se basent davantage sur des "preuves" ou "evidence" en anglais. Elle reflète ainsi la nécessite de renforcer l'interface entre la politique, la pratique, la recherche et l' évaluation.

Cette approche est au cœur de notre Communication sur l'efficacité et l'équité des systèmes européens d'éducation et de formation". Les résultats de la recherche économique et sociale démontrent clairement les bénéfices qu'une relation plus étroite entre les chercheurs et les décideurs peut apporter aux initiatives politiques, favorisant à la fois leur créativité et leur crédibilité. Afin de rendre leurs systèmes plus efficaces et équitables, les États membres devraient pouvoir mieux comprendre "ce qui marche".

Des politiques et des pratiques éducatives basées sur des "preuves"

Tout ceci doit influencer les politiques et les pratiques éducatives.

Trop souvent, les réformes s`appuient sur les perceptions et les croyances du public, les positions et les actions des partenaires sociaux. Pour des raisons de budget, le volume de recherches en la matière est nettement insuffisant. De plus, les Etats Membres accordent un rôle très inégal à la recherche et à l'évaluation dans la détermination de leurs politiques et pratiques

Il faudra, donc, se pencher aujourd'hui sur les trois dimensions du "continuum du savoir" : sa création, sa diffusion et son application.

De nombreuses questions qui constituent la trame de cette Conférence restent ouvertes:

1. D'abord, la création du savoir: quelle politique de recherche éducative mettre en œuvre afin d'accroître l'impact, la qualité et la pertinence des travaux de recherche tout en préservant la liberté scientifique des chercheurs?
2. Ensuite, la diffusion du savoir: comment construire une interface du savoir entre chercheurs, décideurs et praticiens qui permette d'élaborer un transfert de connaissances régulier?
3. Enfin, l'application du savoir: comment promouvoir l'application du savoir par les décideurs et les praticiens tout en encourageant le développement d'une culture de l'évaluation?

La contribution de la commission européenne

Depuis des années, la Commission européenne est devenue un partenaire majeur dans le monde de l'éducation. Aussi, nul ne s'étonnera que la Commission ait d'ores et déjà largement contribué à ce débat. En effet, plusieurs études lancées par la Commission ont analysé le fonctionnement des systèmes d'éducation et de formation, ainsi que les relations entre ces systèmes, la croissance économique et la cohésion sociale. Nous avons également créé des réseaux d'experts pour nous aider dans nos travaux.

Certes, nous ne sommes pas les seuls. Depuis quelque temps, nous observons aussi que le souci d'asseoir davantage la politique et la pratique sur des preuves devient une priorité dans certains Etats. Tout d'abord, il y a le travail de l'OCDE sur ce thème qui vous sera présente après mon intervention. Ensuite, plusieurs Etats Membres participeront en mai à une "activité d'apprentissage par les pairs" (PLA) aux Pays-Bas sur le même sujet. En outre, 6 États Membres ont créé un groupe informel pour renforcer la coopération au niveau européen sur ces questions.

Enfin, la Commission a prévu d'adopter en juillet une Communication sur ce thème qui proposera un plan d'actions visant à examiner dans quelle mesure l'interface du savoir et le continuum du savoir peuvent être améliorés. Le but ultime est de tirer des enseignements aux niveaux local, national et européen pour promouvoir la réflexion stratégique et identifier les priorités futures.

Conclusion

En introduction, je me suis permise d'évoquer devant vous certains chantiers de l'année 2007, de signes positifs pour les politiques d'éducation.

Dans ce contexte porteur, la Conférence d'aujourd'hui peut apporter beaucoup. Elle doit permettre de clarifier les cadres conceptuels et politiques pour les initiatives futures dans ce domaine, en vue notamment de la prochaine Communication de la Commission. Elle peut également fournir un forum pour les résultats. Elle peut enfin faire émerger des propositions prometteuses pour le développement de politiques et de pratiques davantage basées sur des preuves

Je suis convaincue que cette Conférence sera très productive. Elle aidera à établir les bases d'une coopération fructueuse au niveau européen pour renforcer le triangle de la connaissance – éducation, recherche et innovation – qui doit être à la base de taus nos travaux futurs.

Influences of Recent Research on the Education Policy of the Council of Europe: the case of Education for Democratic Citizenship and Human Rights

Olöf Olafsdottir, Head of Department, School and Out-of-School Education, Directorate of Education, DG IV, Council of Europe

Ladies and gentlemen,

I feel honoured to be able to address this very important conference on behalf of the Council of Europe. My thanks go to the German Presidency and the organisers for having invited me, but also for having invited all the member states of the Council of Europe.

Knowing before acting is as essential in education policies as in any other policies. It can be more difficult in education as educational results are not always easy to measure. In particular, it is difficult to obtain reliable knowledge about what is needed in some of the softer subjects dealt with in the framework of education. This is especially true for the work of the Council of Europe where the main fields of activities are education for citizenship and human rights education, history teaching and intercultural education. Over the last few years, the focus has increasingly been on teacher education in all those fields, and this is a direct result of evaluation, research carried out and evidence based feedback from the member states of the Council of Europe.

The Council of Europe was founded in 1949. Pluralism and democracy, the protection and promotion of human rights and the rule of law are the core values of this oldest European organisation. The Council of Europe is a human rights watchdog for Europe's citizens and a guardian of democracy and the rule of law. This function is not only exercised through the monitoring of its human rights instruments, such as the European Convention on Human Rights but also through information, education and training. The two approaches go together and are complementary for building a democratic culture in Europe.

The purpose of my intervention today is to give you an idea of how the Council of Europe's activities in the field of education are influenced – to a considerable extent – by research and evaluation. I will take as an example one of our flagship projects, "Education for Democratic Citizenship and Human Rights" which has been ongoing since 1997.

Over the last fifteen years, since the accession process of the new democracies to the Council of Europe finished, the Organisation has intensified its work on democracy learning. This is based on the conviction of policy-makers that the health and stability of democracy depends to a large extent on the civic engagement and the capacity of European citizens to participate in public life and in society; and that they are not born with these capacities but need to learn them through teaching and practice.

The project, Education for Democratic Citizenship (EDC), launched in 1997, was seen as a public policy issue, as a means of combating political apathy and consolidating democracy in European societies. At the time when it was launched, the situation regarding its implementation in member states differed considerably, but nevertheless it quickly became a priority of education policies across Europe, regardless of the education systems or the cultural and political specificity of the different countries. The Project has had two phases, from 1997 to 2000 and from 2001 to 2004, followed by the 2005 European Year of

Citizenship through Education. It has now entered its third phase which is entitled: Education for Democratic Citizenship and Human Rights.

The reason for the increased focus on human rights is the fact that governments and other decision-makers have also acknowledged the crucial role of education in helping to tackle major societal problems, such as the disintegration of the social fabric, increasing violence, the threat of terrorism, migration within and across countries as well as environmental threats. EDC/HRE does not only teach how society functions but also promotes active and concrete participation in society at an early age, whilst emphasising that individual freedoms are based on solidarity, mutual respect and dialogue.

The first two phases of the Project

The EDC project started by exploring and developing concepts, definitions and strategies. Several texts were adopted at a very high level, emphasising the pivotal role of EDC in education policies and reforms. Among them should be mentioned:-

1) Resolution adopted by the Standing Conference of the Ministers of Education (Cracow, 15-17 October 2000) and
2) Recommendation (2002) 12 of the Committee of Ministers of the Council of Europe to member states on education for democratic citizenship (16 October 2002).

This Recommendation stated inter alia, that EDC should be made "a priority objective of educational policy-making and reforms". It is still the only text adopted at the European level on the subject and has been translated into all languages of the European Union. It was expected that a subject which benefited from such strong political support would soon find its way into the education systems in the member states where it was not yet being implemented.

The All-European study on EDC Policies

However, it also became clear after three or four years, that political declarations were not always followed by practice. It was then decided (2002) to embark on a research project to map out the national EDC policies in Europe, and undertake an All-European Study on education for democratic citizenship in order to better understand the situation and provide an evidence proven base about policies and practices. Seven researchers undertook the work on clusters of countries – Northern Europe, Western Europe, Southern Europe, Central Europe and Eastern Europe. They were assisted by the co-ordinators of the project in all member states. The results were published in 2003 in separate studies for each cluster of countries, along with a synthesis of these studies.

The results can be summarised in the following way:

- Although all countries claimed that EDC was a priority goal, in most of them actual curricular provisions proved to be insufficient. EDC had a weak position in relation to other subjects which were important for final exams or entrance exams, and was not always in the mandatory curriculum. There was little time given to EDC, perhaps 1-2 hours a week, its identity was uncertain and it had little visibility.
- There was an overall pattern of limited, sporadic teacher training related to EDC, with the majority of it generalist in initial teacher education and optional in terms of in service training. This was in clear contradiction with the crucial role of teachers when it comes to democracy learning and developing practice in this field. This finding raised serious questions about the ability and effectiveness of teachers to promote the more active, participatory approaches associated with reforms of citizenship education in many countries. The study noted that "the most powerful lessons that

teachers and schools teach their pupils arise from the way they act and behave, not from what they tell them."

- As to school governance and school functioning, it was observed that in many countries, individual rights, empowerment of pupils/students and inclusive decision-making processes were neither well respected in practice in classrooms nor in schools , nor in the school systems themselves. This was perhaps one of the most difficult things to deal with, given that in many countries there was – and still is – a tradition of the authoritarian way of school governance.

Other important findings were the following:

- A tension was observed between the main actors in the public arena. Government structures were seen to have limited ability to implement decisions regarding EDC/HRE. There were often other more pressing policies dealt with both at the administrative level and in schools, even if those policies lacked political support. For those in charge of the implementation within the public administration, EDC/HRE sometimes seemed an abstract and distant goal, and not within their immediate responsibilities. They preferred to work on other subjects, more familiar to them.
- A positive trend observed were the numerous initiatives taken by civil society (NGOs) to promote EDC/HRE. However, there appeared not to be enough co-operation between these initiatives and the top-down governmental processes. At the same time, it was clear that for EDC policy to be successful it is essential to mobilise the capacity and expertise of various stakeholders. Only in some countries (basically the Nordic countries and some countries of Western Europe) does the state renounce its monopoly as the sole provider of educational services and shares with partners from civil society, especially when it comes to adult education.
- The final finding which should be mentioned here was that monitoring, evaluation and quality assurance are essential to successful EDC/HRE policy implementation. This indicated once again a difficulty for the implementation of this type of education, given that its assessment of knowledge is not easy.

What were the main challenges that came out of this study and how was the policy of the Council of Europe regarding education for democratic citizenship and human rights adapted to these results?

In many of the member states, they were the following (this information comes mostly from the synthesis of the regional reports):

- Reducing the compliance gap between EDC policies and practices;
- Improving and extending the participation of students and community representatives in the education system, particularly in school management;
- Developing more effective and comprehensive teacher training, at both pre and in service levels;
- Introducing a culture of and suitable measures for monitoring, quality assurance and evaluation;
- Agreeing on and fighting for the place of EDC within competing educational forms and priorities.
- Strong focus was therefore put on awareness-raising in EDC/HRE – for instance through the organisation of the European Year of Citizenship through Education (2005) as well as on the development of practical tools to assist practitioners with implementation.

Once the "Year" was finished, it was decided to work on three strands: policy development, teacher training and democratic school governance, where;

- In *policy development*, the emphasis is now on building sustainable policies in EDC/HRE through development of curricula, criteria for competencies and tools for quality assurance; through the development of effective practices, through building a proven base of what works and what does not work.
- In *teacher education and training* a significant effort is being made to increase the qualifications and professionalism of teachers and educational staff, through the definition of core competencies for the teaching profession, preparation of methodological and educational resources;
- Regarding the *democratic governance of schools*, examples of innovative/successful practice are developed and shared, and tools have been prepared for the use of head-teachers and other educational staff.

A number of tools have been prepared and others are in preparation (quality assurance, democratic school governance, teacher training, tool for policy makers, manuals for schools). Some of them have been tested in schools, whilst others will be tested within the coming months.

In line with the importance of networking and co-operation, the Council of Europe works closely with several other institutions and organisations, such as the European Commission on indicators for measuring active citizenship and education for active citizenship; with UNESCO and the UN High Commissioner for Human Rights on all the subjects mentioned; with NGOs and Foundations on the questions of quality assurance and democratic governance of schools.

Bilateral activities are also organised, especially in the newer member states. They range from preparation of EDC/HRE programmes, plans and school curricula, to teacher training and work on self-development of schools.

For the moment, it is foreseen to continue the EDC/HRE project until the end of 2009. However, given the nature of the subject and its close relationship with the priority issues of the Council of Europe, that is democracy, human rights and the rule of law, it might well become an ongoing project of the Organisation. In any case, before the end of 2009, new research on the situation will certainly be required, and its results may once again change the focus of our work.

OECD and Evidence-Informed Policy Research

Tom Schuller[2]*, CERI/OECD*

Introduction

In one sense, it is curious that we should now be addressing the issue of evidence and educational policy, since arguably all the work of OECD is about producing evidence to inform policy. OECD is not a legislative organisation (though it occasionally produces so-called "soft law", ie guidelines and the like). It concentrates its efforts on producing statistics and high quality analysis, and its main clients are governments (at national but also regional/state level) and other policy-makers. However there has been an undeniable upsurge of interest in OECD member countries in the issue of how evidence to inform educational policy is produced and used, and how these processes might be improved. In this paper I give some background to relevant OECD work on this issue; indicate the key themes we have been pursuing; and conclude by identifying some of the key challenges.

Background

In 1995 OECD published a report on *Educational Research and Development: Trends, Issues and Developments*. This pointed to a perceived weakness in the production of educational research, and a patchy state of knowledge about education systems. The years around that date saw huge steps forward in the OECD's own production of data on education, with the rise of educational statistics demonstrated in the physical expansion of *Education at a Glance* from a slim booklet (matching its name – ie it could be looked through very quickly) to a large volume running to several hundred pages. At the same time PISA began its phenomenal rise, based on the direct cross-national measurement of educational outcomes, as opposed to the use of proxies such as qualifications. Other international agencies made similar efforts, either at the statistical level (Unesco) or with comparative analyses such as the International Education Association's studies of mathematics or civics.

Through CERI, OECD continued its reflections on the state of the knowledge base via a number of studies and publications on knowledge management. These looked at how knowledge is produced and utilised, not only in education but in other sectors, notably health and business. Some of this work was aimed at developing models for analysing knowledge production (see OECD 2004 for the "four pumps" model which attracted a lot of interest). The general conclusion was that education, as a major service to which a large proportion of public expenditure was devoted, had a weak knowledge base and was less than innovative in the extent to which it drew on research to change its delivery.

Running in tandem was a series of national reviews of educational R&D. The series began with New Zealand in 2001, and carried on through England, Mexico and Denmark to conclude with Switzerland in 2006. The Danish review saw the production of a "template" which could be used generally to promote information and debate on the state of play of educational research – itself a significant output from the work. Each country had of course its own specific characteristics, and strengths and weaknesses. However there were some fairly clear common features, which can be summarised as follows.

---

2 Tom Schuller is head of the Centre for Educational Research and Innovation at OECD. This chapter is written in a personal capacity and does not engage the OECD.

First, by comparison with the sums spent on education as a service, investment in R&D was very low. Probably the most suitable comparator is the health sector, as a very large-scale predominantly publicly-funded service, and a reasonable estimate would be that something like five times as much is spent proportionately on R&D in health as is in education. (The exact comparison depends in part on how much weight is given to expenditure on R&D for drugs, but the overall picture is clear).

Secondly, the education research capacity in most countries is quite low. Defining this capacity is itself a problem, since education potentially spans many boundaries and disciplines, and university education departments are only one source of relevant research. But the general picture is of a research field where research is often quite fragmented, low-rated compared with other disciplines (including other applied disciplines) and with a low level of methodological sophistication. Two particular concerns were the lack of cumulative knowledge, ie the fact that there was not a common core of research knowledge which accumulated steadily over time; and the lack of effective utilisation of large-scale datasets, and a related lack of methodological capacity.

Thirdly, the links between research and policy were seen as inadequate to ensure that research had an effective impact on policy. This conclusion did not necessarily imply that all the fault here lay with researchers; it was a perception that the interactions between the research and the policy communities were often weak.

Against this background, there were a number of drivers behind the renewed concern for improved evidence in educational policy:

1. An overall concern with education as a key component in innovation and growth. The rhetoric about knowledge societies is sometimes overblown, but the shift to a more service-oriented and technology-biased economy undoubtedly places greater stress on the competences of countries' populations.
2. Understanding how effective learning happens, and how good educational systems and institutions are at delivering effective learning, therefore forms part of a wider concern with accountability. Pressures on public expenditure are almost universal, and increasingly politicians and the people they represent have come to demand more in the way of information and analysis on how well these expenditures are being used.
3. The general pressure for greater accountability takes more specific forms of concern for quality and for effectiveness. To the extent that educational research has often not been able to demonstrate a clear-cut contribution to improvement in these respects, questions began increasingly to be asked not only about the education system per se but about the role of educational research as a source of progress in education.
4. Public dissatisfaction and its expression have been given a greatly enhanced profile by new technologies, notably the Web. This has both given people greater access to information – for instance to test scores or league tables but also to information which is sometimes of dubious quality – and raised their expectations of what will be provided in the way of knowledge. At the same time, it has increased hugely the information to which policy-makers have access, but in some senses has thereby made their tasks more not less difficult, as quality and selection become more urgent matters.

In response to these trends and concerns, OECD took a further step with the launch of a series of four international seminars. The themes for these seminars developed on a rolling basis. The first took place in Washington DC, in March 2004. It focused very much on experimental design in educational research, and even more specifically on the use of randomised controlled trials as a – or the – key means of producing robust causal evidence. This was a controversial theme which provoked significant differences of national position, with interestingly fruitful consequences. Notably some participants who were quite

strenuously opposed to RCTs nevertheless were brought to acknowledge that their own research systems were less than adequate when it came to producing knowledge on which key policy decisions could be based. The second seminar, in Stockholm in January 2005, took a much broader perspective, looking at the interactions between the main stakeholders in the policy-research game.

The third seminar, in The Hague in September 2005, swung back to a more defined focus, but building on the theme of interaction between the stakeholders. This seminar saw the emergence of a new line of thinking, namely the role of so-called brokerage agencies (see below). The key issue here was to ask what the experience had been to date of those countries which had set up mechanisms aimed at improving the communication between research, policy and practice (it was recognised at this meeting that practitioners were an important player in this nexus). A final meeting, in London in July 2006, brought together some of the conclusions from the series, but also dealt with the issue of implementation and scaling-up: how can we proceed from research results which may come from only a limited context or sample in order to influence the field as a whole.[3]

The inputs and discussions from these seminars have been brought together in a recent OECD publication, *Evidence and Education: Linking Research and Policy* (OECD 2007). Annex A gives the contents of this publication. I turn now to deal with some of the specific issues arising: in the next section, with internal capacity issues within the research community, and with aspects of the interactions amongst stakeholders in the research-policy debate; and in the following section with brokerage.

Methodologies and capacities

Quite rightly, there is considerable variation in the types of methodology used to conduct educational research, including secondary analysis of large-scale datasets, qualitative and case-study investigation, discourse analysis of policy positions and experimental investigation. Quite rightly too, there is disagreement over the relative status and application of these different methodologies. It is entirely natural that there should be divergences of view. However the nature of these methodological divergences is not always helpful.

All parties would agree that rigour is essential. I would not dream of attempting to adjudicate between the rival claims to rigour. But whatever the definition is of rigour in the theoretical sense, one key point at issue is what the appropriate balance is between rigour in some absolute sense and the need for pragmatic accommodation to the limitations of time and resources. In other words, ideal forms of scientific inquiry may come sharply up against a number of constraints, such as limited budgets for research; the need for quick decision-making on the part of policy-makers; or even the inherent complexity and instability of education systems, so that even long-term, large-scale and multi-method research cannot produce definitive results. This tension between scientific ideals and 'best available' pragmatism is inherent in the field.

However, to be a little provocative we could identity three features which, in an ideal-type sense, distinguish debate in different contexts or countries:

a.) *Warfare*. Here, parties holding to different positions exhibit latent or overt hostility towards each others' positions (though they may or may not express themselves with perfect courtesy). They are simply unwilling or unable to recognise the merits of others' approaches; the paradigms are too much in conflict with each other.

b.) *Mutual invisibility*. Here there is no friction because there is no contact (and therefore, we may add, no light). Parties ignore each other, make no reference to each others' work and do not engage in dialogue but pursue their own lines of discourse.

3 For detail on these, see www.oecd.org/edu/rd

c.) *Complementarity*. Here the parties recognise that there is no single best way, that there are different methodologies which are appropriate in different contexts, and seek to identify the best mix of approaches for any given issue or problem. The complementarity occurs over time rather than cross-sectionally, so that different approaches help each other to test and refine the knowledge base on an ongoing basis.

Clearly these three options are not value-free, and I should emphasise that they are ideal types and not representations of any actual situations. Almost every country will contain, in different corners of the educational research community, elements of all of these. The typology is there for people to reflect on in application to their own specific contexts.

I turn now to capacities. For the purposes of this paper, I offer simply two dimensions of this issue.

a.) *Deepening/broadening*. In some cases the priority need is for deepening certain capacities, ie where a country simply does not have enough people with expertise in a given form of research. As mentioned above, the OECD R&D reviews identified a lack of capacity, and in some cases referred specifically to the lack of expertise in handling large-scale datasets. This needs further empirical verification. An alternative judgment could be that the research community (as opposed to individual researchers) needs to expand the range of expertise at its command, because there is over-reliance on a narrow set of techniques. Of course both deepening and broadening may be desirable, but there may still be interesting discussions on what the capacity priorities are.

b.) *Producer/consumer*. Most of the discussion on capacity is framed in terms of producer capacity, that is the collective competences of the research community. However if we recognise that effective research-policy links entail two-way communication, this implies that we should not ignore capacity issues in relation to consumers of research – in this case primarily policy-makers, but also practitioners. Understanding enough about research approaches and processes is important if policy-makers are to be able not only to understand results but also to specify properly the research which they wish to commission or otherwise support.

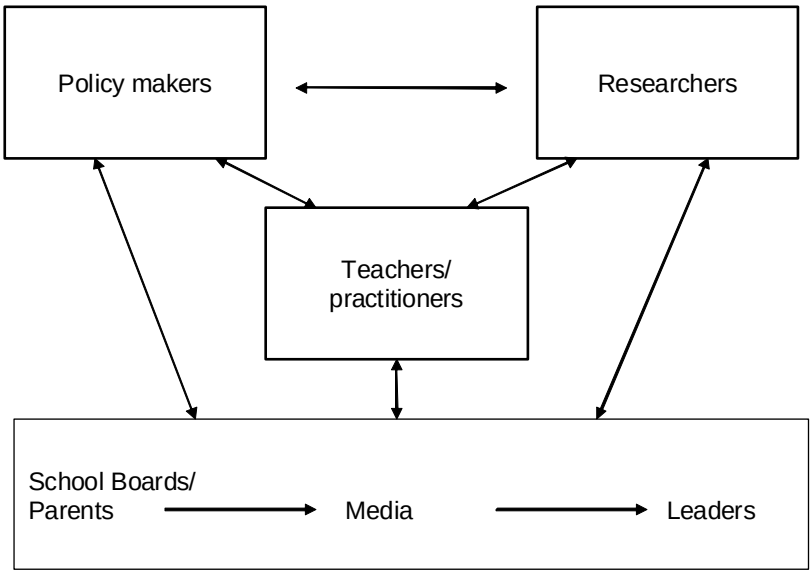
Both these dimensions therefore raise at least two leading questions;

1. What forms of capacity are most in need of strengthening?
2. How and by whom should this be done?

Stakeholder networks

Figure 1 gives a very simple outline of links between some of the major stakeholders in the research-policy debate. We have highlighted policy-makers, researchers and practitioners because of the particular focus of this OECD activity, but others obviously have a part to play. For example, the media influence the mature of research and of political responses, for example by seizing in a hostile fashion on negative results, or oversimplifying outcomes to achieve a headline. More positively, good media reporting raises the public understanding of an issue.

Primary lines of communication/interaction



The primary value of this simple diagram is heuristic. In any given context it enables us to ask questions of the order:

- What are the strengths and weaknesses of each of these links?
- What is the quality of the information flows?
- What could be done to improve each of these links?

Brokerage

As noted above, one of the key new issues to emerge from the discussions was the potential importance of brokerage. Brokerage is understood as the processes by which information is mediated between stakeholders; the processes include formal and informal mechanisms, and in some instances agencies specifically set up to carry out this function. Examples of brokerage described in *Evidence in Education* include the EPPI Centre in the UK (Evidence for Policy and Practice Information); the What Works Clearinghouse in the US; the Iterative Best Practice Syntheses in New Zealand, and the Canadian Council on Learning. Interestingly, the renewed concern with evidence-informed policy making has already spawned two new brokerage initiatives, the Knowledge Chamber in the Netherlands and the Knowledge Clearinghouse in Denmark.

The emergence of these agencies throws up a number of issues surrounding their functions:

1. *Dissemination*. In some cases, the goal of brokerage is to mobilise different forms of dissemination, in order to make evidence and research results more easily available to those interested. This entails writing of summaries in accessible language, but also the use of the internet and roadshows in order to achieve a much wider audience than standard publications would.
2. *Interactivity*. A step beyond one-way dissemination is the promotion of interactive debate on the results. This might be through conventional means (newsletters etc) or through web-based discussion groups and other more innovative means. The New

Zealand initiative puts particular emphasis on an iterative and participatory process of continual refinement and development.

3. A very different goal is the establishment and enforcement of standards of *rigour and validity*. Thus the WWC in the US has transparent and strongly defined standards on what should count as valid research results, and confines its dissemination efforts to those studies whose evidence conforms to those standards.
4. Even more broadly, agencies may seek actively to build *trust and cooperation* between different partners, as part of the goal of improving the use of evidence.

CERI is currently embarking on a comparative study of different brokerage agencies. The study will focus on these different functions, and how they are interpreted in different countries. It will also seek to establish how far the agencies have made an impact – in other words, it will look for evidence about evidence-use.[4]

Future challenges

There is little doubt that the pressure for moving further towards evidence-informed policy – making will increase. Debates on what counts as evidence will continue; it is to be hoped only that they will be constructive rather than polarising. In conclusion, here are four challenges which are likely to shape the future contours of the debate in OECD countries and beyond.

1. What progress can be made towards an "evaluation culture"? By evaluation culture I do not mean simply a context where many evaluations take place, however high quality these may be. It is a culture where individuals and institutions are gearing to seeking and using evidence about performance and effectiveness, in ways which have an impact but are not punitive. Accountability demands that assessments are made, but these are not one-off occurrences but part of an ongoing process of feedback and development.
2. How do we identify key capacities for producing and using good evidence? The discussion above addressed some of the issues in this respect. The answers will not be the same across all countries, but addressing similar questions could lead to particular useful comparative information.
3. How to promote mixed-methods approaches in order to generate more rounded evidence? It is easy to plead for using different methodologies, but much harder to find solid examples of mixed-method approaches which successfully capture different angles by producing different kinds of evidence. The challenge here is not so much to aim for research programmes which are themselves fully integrated, as to develop processes which over time allow different methodologies to be brought to bear on a given issue, so that over time evidence is tested and refined by findings from different angles.
4. Finally, how to build social capital within research? By social capital I mean, in a standard way, the norms and networks which enable greater collective results. This is not incompatible with competitive research, which is a necessity for quality and innovation. It means seeking to combine this competitiveness with the trust and cooperation which reduces transaction and monitoring costs and which, above all, allows evidence to be continually improved, built upon and used for action.

Success in answering these questions would take us a long way towards changing the situation we described in the OECD reviews and earlier analyses. Our perception is that

4 For further information on this study, contact tracey.burns@oecd.org

there is already a good deal of forward movement, and we look forward to seeing the evidence that this is being sustained and put to good effect.

References

OECD (2004), Innovation in the Knowledge Economy, OECD, Paris

OECD (2007), Evidence in Education: Linking Research and Policy, OECD, Paris

Annex A

Evidence in Education: Linking Research and Policy

TABLE OF CONTENTS

Knowledge for Action in Education Research and Policy: What we know, what we don't know and what we need to do

Ben Levin, Ontario Ministry of Education/Ontario Institute for Studies In Education

Introduction

This paper represents the written version of a presentation given on Wednesday, March 28, 2007, at the above conference. As such, this paper is not intended to be a full academic treatment of the issues and provides only limited references to related literature. Fuller discussions with more related references can be found in Levin, 2004, 2005, 2006.

The views in this paper are solely those of the author and do not necessarily represent the policy or opinion of any other person or organization.

A note on language

One problem with the whole field of research-policy links is the lack of an appropriate language (at least in English). Terms such as "research impact", "research use" or "dissemination" imply a one-directional relationship from research to policy. However the relationship is actually much more complex, as illustrated later in this paper. We simply do not yet have – at least in English – the right terms to describe this more complex relationship.

The knowledge base on research impact

Knowledge about the impact of research on policy, while growing, remains limited and fragmented.

Conceptual and empirical work on the issue is being done in many different settings, with health playing a leading role but many other fields such as criminology or social welfare also contributing. Relevant ideas and evidence can also be found in political studies (such as work on the role of research in political agenda setting), in cognitive psychology (for example on understanding how "facts" come to be salient for people), in social psychology (around group knowledge and decision processes) and in organization and management theory. However the various efforts across fields are poorly linked, with researchers often being unaware of high relevant work in other disciplines (an interesting example of the challenges of mobilizing knowledge for action!). Also, as argued later, there is still a substantial need for both better conceptualizations and more empirical evidence on research-policy relationships.

What we know

There are many indications that research and empirical evidence are playing a greater role in shaping actions and policies of governments. There is more attention to research in the media; it would be unusual in Canada to find a day in which there was not some media coverage of research, be it from health, education, social policy, the environment, or other fields. Increased media attention is both triggered by and leads to increased public interest, which in turn triggers growing government interest. At a time of high cynicism about politics, governments also see research as providing legitimacy to their policies. While this

interest in research is the case in all areas of policy, it is perhaps a greater change in education, which has less tradition than some other policy fields of using research to shape policy.

The same could be said about the education profession itself. In most countries until quite recently many professional educators had little regard for the value of research – in part because the body of research did not provide them with much useful guidance and in part because of a belief in teaching as a craft not amenable to evidence or generalization. These views have changed in many countries as teachers are better educated, as they have more exposure to research, and as the body of relevant education research has increased in quantity and quality. So in addition to a growing public and political appetite for research findings, there is also strong professional interest.

One should not underestimate the significance in this change of the growing body of knowledge emerging from education research. The research-policy relationship is one of both supply and demand – more knowledge yields more interest in knowledge, just as greater interest is likely gradually to yield greater knowledge. There is now enough good evidence in a number of areas to guide policy and practice. To take just a few examples:

- smaller classes are linked to modest gains in achievement in the early grades;
- retaining students in grade is an ineffective practice;
- early reading development benefits from explicit teaching combined with exposure to a literacy-rich environment;
- increased high-stakes testing will lead to an increase in avoidance or cheating strategies;
- changes in school governance are at best weakly related to student outcomes;

And so on. While these findings do not always shape either policy or practice yet, they are a lively part of the debate on all these issues.

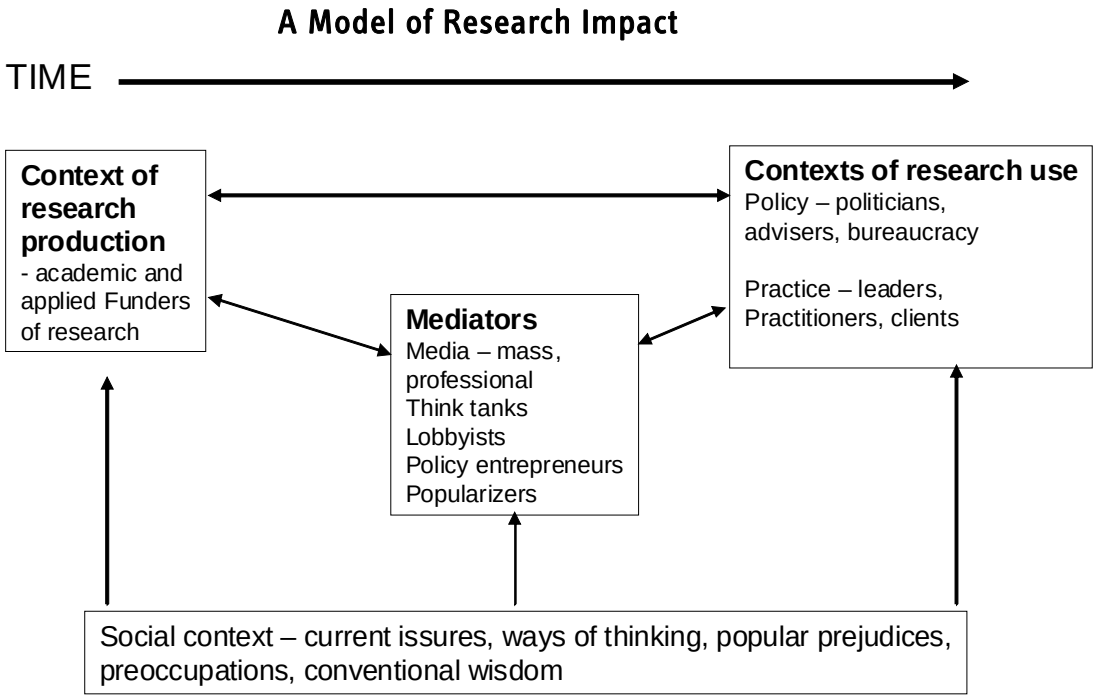
While interest in research is growing, it remains rare for there to be a direct link between a particular study, or even a set of studies, and education policy. In education, as in many other fields, the impact of research is primarily indirect and gradual.

Models of research-policy impact

Many researchers assume a situation in which the results of research are immediately taken up and applied by policy-makers, a kind of direct transmission model. That is clearly not the case except perhaps in highly unusual circumstances. Rather, the research-policy relationship is something closer to the following diagram, though this, too, is only a partial representation.

Several key points can be made about this model.

First, there are a variety of links, some stronger and some weaker, between research and policy (with the policy world having its own bifurcation between the political and civil service elements).



Second, the impact of research happens largely through other social and political processes, and is therefore mediated by the broader social and political context and the climate of ideas in society.

Third, many of the links from research to policy are indirect or mediated through a variety of third parties.

Fourth, the capacity of various "users" – in this discussion policy-makers – to find, understand and apply research is itself an important issue.

Let me say a little more about each of these points.

Multiple links

Research reaches the policy world through multiple routes. The policy world itself is not unitary, having two major elements of political actors and their civil service staff. These two groups often have quite different and sometimes entirely divergent values and worldviews. For example politicians must live within electoral cycles and respond to political actors to a much greater influence than do civil servants, who may in turn feel more allegiance to their policy sector than to the government of the day (see Levin, 1986 and 2005, for a fuller discussion). Of course some policy makers – politicians and senior civil servants – do read research and talk directly to researchers, even though time constraints and other pressures generally make this difficult. Some researchers or research institutes, such as universities, make very deliberate efforts to draw their work directly to the attention of policy makers.

However such direct connection is the exception. For the most part research is a part of larger social and political processes. For example, research findings are often used by lobby groups arguing for (or against) particular policies. Research is also, as mentioned, often reported by the media and so enters public consciousness and government awareness through that route. Nor is this only or even primarily through the mass media; specialized media such as industry or sector publications often play a critical role in making research more available.

This mediation process means that awareness of research is affected by broader currents of social thought – what Schon (1971) called "ideas in good currency." If people are generally interested in something, whether it is high school graduation rates or global warming or family violence, then research on those issues is more likely to be taken up, especially if the research is consistent with current public opinion (or, sometimes, if it is highly divergent from current opinion). As we all know, issues go in and out of public favour so that a finding that would be of considerable interest at one point may get no attention at all a year later.

This gradual and indirect pattern means that various third parties play a vital and generally underrated role in the process of research impact. A wide variety of third parties be important. The role of the media has already been discussed. Other third parties include political actors such as lobby groups, advocacy groups for particular causes, professional or trade associations, unions, employer groups, and the various bodies that make up civil society. In addition, there is in many places a category of persons or organizations that might be called "policy entrepreneurs" (Mintrom, 2000) – people who make it their work to popularize research findings in support of particular causes or positions. In education these would include some of the top national and international speakers, writers and providers of professional development.

Fourth, the capacity of various "users" – in this discussion policy-makers - to find, understand and apply research is itself an important issue. No matter how good, relevant or plentiful, research will only matter if others are able and want to use it.

Interest in research is shaped by people's context and work. To be used and useful, research must "make sense" to people. This means that the take up of research is greatly affected by the realities of people's work. Knowing that something is a good idea does not mean it will be adopted – witness the fact that many people fail to exercise or continue to smoke despite their knowing that they should do otherwise. Research results that seem unworkable will not be applied no matter how relevant they are. We know from work with professionals, including teachers, that they must see how they can actually apply a new approach before it will be adopted, no matter how promising it may seem. Political or civil service actors may similarly disregard research if it seems too out of step with what they believe the public or the education sector itself would accept.

The implication of this model is that knowledge is social, that it grows from previous ideas and relationships. And this means that increasing research impact will require stronger connections between the world of research and the world of policy, worlds which are very, very different from each other. The world of research tends to be long-term, slow-moving, focused on a small number of issues, and fairly insular. The world of policy is the opposite – fast-moving, with short timelines, many issues operating simultaneously and highly open to a wide range of outside influences.

The real world of politics is very poorly understood by most people, who regard it as irrational and devious. In my experience this is not at all the case. Space does not permit even a brief treatment of all the relevant aspects of the political world (see Levin, 2005), but among the factors that can be mentioned as constraints on political choice are:

- the huge range of issues that have to be considered simultaneously;
- the lack of timely information on many of the issues;
- the enormous pressures exerted by various interests on decisions;
- constant scrutiny and assumption of wrong doing by opposition parties and the media;
- the potentially enormous consequences of even small mistakes.

It is impossible to provide sound advice to political decision-makers without a feel for the formidable challenges faced by all politicians and their staffs.

Increasing Efforts

The growing interest in research and empirical evidence has led to a wide range of efforts to strengthen links between research and policy, in education and other fields. Many different kinds of organizations have been involved in these efforts, including governments, universities, think thanks, special-purpose national agencies, and a range of other organizations. Some examples follow, chosen to illustrate some of the different strategies now in place. These examples come largely from the author's direct knowledge resulting in examples tilted towards English-speaking countries, and particularly Canada. No claim is being made that these examples are the best that could be found; rather they illustrate a range of approaches.

Universities. As the largest single source of research in education, universities have made some efforts to improve policy impact. A few have dedicated websites linked to education research, while others organize specific events to connect with policy-makers. However in general universities have much more elaborate infrastructure to support commercialization in areas of science and technology than they do for research in the social sciences.

Governments. Although governments are increasingly interested in evidence as a basis for policy, their promotional efforts in this area are largely through their funding of research and their support for third parties. Many governments support dedicated research creation and dissemination institutions in education – such as the EPPI-Centre in the UK, or the regional labs in the US, or the OKI in Hungary or the DIPF in Germany. New approaches and models are being created rapidly. Although many governments announce their active interest in research, one finds much less emphasis in this area on the websites of education ministries, few of which have any overt information about or links to research. It is also worth mentioning the efforts of national statistics agencies to go beyond simple data collection to analysis and dissemination though here, as in every other category, the extent and quality of effort is highly variable.

National agencies. As mentioned above, a number of countries have created national agencies or organizations dedicated to some form of research dissemination. Often these agencies are also engaged in conducting research. Sometimes they are also in the business of funding research, or, to put it another way, research funding bodies are showing increasing interest in showing that the work they support does have an impact. External agencies are seen as desirable in that they have fewer constraints than do governments. However there is also the risk that they will not have the strong connections with either government or the schools sector that are required to have real impact. Many different approaches and models seem to exist around the world, creating a great opportunity for more connection and learning across these endeavours about effective practices.

International agencies. Most of the large international agencies give at least some attention to relationships between research and policy. The EU has taken this as an important area of work, although the organization's website still gives little indication of specific actions taken. The OECD also has a mandate around research impact but its practices are still fairly limited. Some parts of the organization, notably CERI, are studying research impact, but this has not yet translated to the way the organization as a whole promotes and shares its work and findings. The same could be said about other large international organizations such as UNESCO or IIEP. The challenges of creating research impact in international organizations are, of course, even more formidable than those in national settings.

Assessment of efforts so far

The above is far from a thorough survey of efforts to link research and policy. Nonetheless, even based on this selective review, some comments can be made. There are many interesting ideas and initiatives being tried in many places. These are highly eclectic, including different forms of reporting, events to bring researchers and policy-makers together, and, particularly, efforts to use electronic communications vehicles such as RSS or podcasting or chatrooms. Many efforts are one-shot but others are sustained.

Overall one seems still to find primarily one-way approaches, starting with research and working out to reach the policy field through better forms of dissemination. The creation of real and lasting two-way linkages is less common – and clearly more difficult. It seems reasonable to say that we are in the early stages of this effort, with much experimentation not necessarily based on clear specification of either means or ends.

The development of so many different efforts to link research and policy cries out for more empirical study and for a better framework to understand the range of practices and the various logics behind them. One attempt to develop a taxonomy has been made by Walter, Nutley & Davies (2003), in Scotland. They reviewed 200 different research impact strategies (related to affecting teaching practice as well as education policy) and categorized them on two dimensions. First, they rated practices by intervention type based on form and content, with about 40 categories ranging from computer support systems to networks to print materials. Second, they defined eight general mechanisms through which the activities operated – dissemination, education, social influence, collaboration with users, incentives, reinforcement, facilitation, and multifaceted. The authors note that this is a first effort, and the categories are primarily heuristic rather than conceptual. However this work is an important step towards better conceptualizations of research-policy-practice linkages, as discussed a little later.

Current weaknesses

Not even the most optimistic proponent of current approaches would argue that we know and are doing all that we need to. Serious weaknesses remain both in the way research is produced and communicated, and in the way it is taken up by educational organizations and policy-makers (these are discussed in more detail in Levin, 2004).

From the standpoint of research "production", it is clear that the total amount of research being done in education remains far too small, especially in relation to the size of the education sector. An OECD report (OECD, 2002) notes that no country spends even 1% of total education public expenditure on research. The research that is done in education tends to be small-scale and has insufficient attention to replication and gradual production of robust knowledge through building carefully on previous work. The bulk of education research is done in universities and is driven by individual researcher interests so is often weakly linked to larger public policy priorities. For example, a great deal of research in education is about initial teacher training, even though many other areas would appear to have stronger impacts on educational outcomes. And, as already noted, relatively few researchers or research organizations are giving significant priority to linking their work to policy and practice.

The problems, weaknesses and limitations of research "use" are less often discussed in the literature than are issues of research production and dissemination, but they are just as important. Very few educational organizations, from ministries to regional bodies to schools, have an organized capacity to find, distribute and use relevant education research. Even units dedicated to research in ministries of education, where they exist, tend to get absorbed in ancillary work such as issues management, briefing materials for ministers

and senior officials, or analysis of internal data. At more local levels there is even less capacity. Dealing with research tends to be low on the priority list, and one of the first things to be cut when budgets are tight. Thus even if we had wonderful, highly relevant, clearly stated research results available, the capacity to make use of them is in short supply.

What We Do Not – and Should – Know

The problems we face go beyond organization and resources. One reason current efforts are sporadic is that the conceptual or theoretical basis for discussing issues of research impact remains weak. There are few conceptual models available to guide our thinking or practice. As already noted, we lack even good terminology for what we wish to do. Implicit in the above discussion is the idea that there is widespread agreement on what we mean by "research" and what we mean by "use" or "impact", yet even a brief discussion among anyone involved will show that these ideas are themselves contested, with many different ideas about what should count as research for purposes of affecting policy, and just as many ideas on what sorts of connections can or should exist between research and policy. Further careful thinking and conceptualization is urgently needed, even as further efforts are made to create improvement. In my view efforts to define phenomena and efforts to "do" them can and often should proceed simultaneously, with practice helping to define theory and vice versa.

One area that will need particularly careful thinking has to do with the potential connections between the worlds of research and policy, different as they are. Are there some possibilities that seem, in principle, to be more likely than others? Where are the potential areas of overlap in interests, ideas and people?

Gathering more empirical evidence on the effects of various strategies for connecting research and policy is a second major area for more work. As already noted, there are many interesting efforts being made, but very little good evidence on their value. For example, almost every organization is putting resources, sometimes substantial, into improved websites for sharing research. These sites use a variety of tools. But there is little if any empirical basis for determining which tools reach which audiences, or for assessing the relative cost-effectiveness of the different strategies, or even to help us know if sharing through the web is a better strategy than promoting more face to face encounters. Is it more effective to give away material such as research reports than to sell it? Do more people actually use executive summaries? Does providing a good summary encourage people to read the full report, or simply to draw premature conclusions? Gathering good evidence on these questions is challenging, but no more so than many other areas of social science. In some ways the task is akin to research on marketing, so that is another interdisciplinary connection to be made.

More empirical work is also needed on the work of third parties and research brokers. The political studies literature could be useful in helping us understand the way in which ideas move into the political and policy world. Who are the various actors and what roles do they play? Might convincing a few key "retailers" of ideas have more impact on teacher practice than providing high quality resource materials directly to schools?

What we Should Do

Following from the above, three kinds of further action are needed – better thinking about the phenomena in question, and more evidence on the results of various efforts. Of course these are related, in that conceptualization informs the gathering of evidence and evidence helps shape better conceptualization.

This work should involve both international and interdisciplinary efforts, and in particular the development of both research and implementation networks at national and supranational levels. Canada has had good success with funding networks of researchers who are physically widely separated but share common interests; over time these networks can become very effective vehicles for developing high quality work that can reach many different research institutions. Some networks have had good success in including non-academic partners, though this does not always work and always takes extra effort and skill. The EU already funds international and interdisciplinary research, so could readily support networks and projects in the area of research impact.

The field of research impact would be a good candidate for some true experiments as well. It would be possible, for example, to have a website that randomly distributed various products or contacts to visitors; with enough participating organizations a picture of real effects could emerge. (This idea obviously requires more development than is offered here; I put it forward only to indicate the potential range of research methods that could be employed and the quality of work that could be generated.)

It is not only research on research impact that is needed, however; it is even more to create the learning mechanisms in this area. This German Ministry/EU conference represents an important effort to generate interest and activity, and to do so by bringing together a wide variety of people. It must, however, be more than a one-time effort. In my opinion face to face interaction among people remains one of the most powerful vehicles for learning and for building the connections and trust that enable other modes, such as technology-led networks, to be successful. So an important goal for all parties should be to create ongoing opportunities for mutual learning as well as joint work. These learning opportunities obviously need to involve policy makers – both political and civil service – and practitioners, as well as the third parties – brokers, media, lobbyists, civil society – that play such a critical role. Exchanges among the various parties, whether at regional, national or international levels have the potential to be highly stimulating and valuable in advancing our understanding and improving our efforts.

References

Levin, B. (1986). Uneasy bedfellows: Politics and programs in the operation of government. Optimum 17(1), 34-44.

Levin, B. (2004, October 17). Making research matter more. Education Policy

Analysis Archives, 12(56). Retrieved from http://epaa.asu.edu/epaa/v12n56/.

Levin, B. (2005). Improving Research-Policy Relationships: The Case of Literacy. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood and D. Livingstone (Eds.) International Handbook of Educational Policy. Dordrecht: Kluwer. 613-628.

Levin, B. (2005). Governing education. Toronto: University of Toronto Press.

Levin, B. (2006). How can research in education contribute to policy? Review of Australian Research in Education 6, 147-157.

Mintrom, M. (2000). Policy entrepreneurs and school choice. Washington: Georgetown University Press.

OECD (2002a). Educational research and development in England: Examiners' report.

Paris: OECD. CERI/CD(2002)10

Schon, D. (1971). Beyond the stable state. New York: Norton.

Walter, I., Nutley, S. & Davies, H. (2003). Developing a taxonomy of interventions used to increase the impact of research. Unpublished discussion paper, Research Unit for Research Utilisation, Department of Management, University of St. Andrews, Scotland. Available at: http://www.st-andrews.ac.uk/~cppm/Taxonomy%20development%20paper%20070103.pdf

National Experience Report: The Danish Case – Institutional concentration

René Bugge Bertramsen, Deputy Director General, Ministry of Science, Technology and Innovation, Denmark

Introduction

This article will explain how Denmark has aimed to build a knowledge basis for evidence-based education policy. The Danish case illustrates that to establish a coherent infrastructure for educational Research & Development a constant willingness to reform the institutional structure is required.

Since 1999 educational R&D has been a top priority for both the former and the current Danish government. Especially the OECD's PISA results in 2000 and 2003 were an eye opener for decision makers at all levels. The PISA examinations indicated that the competences of the Danish pupils at the age of 15 in general were close to the OECD average. To a small, open economy - which depends on the ability to integrate a large share of the population in knowledge based professions - the results have been disappointing and testify to the need to strengthen the quality of education.

Educational R&D is seen as an effective means to strengthen the quality of education and addressing the problem of weak performance - not only in the primary and secondary schools but at all levels of the educational system. The effect is that children perform better in reading, math and science and go on to either vocational or higher education. A focus area for the government is that at least 50 per cent of young people should complete a higher education programme.

For the current as well as the former government it has been crucial to establish a coherent institutional infrastructure for educational R&D. The central aim is to ensure the production of relevant research and to ensure the access to recent research-based knowledge used in education by teachers at all educational levels.

Multiple actors undertake educational R&D in Denmark. However, the usual perception is that the 8 Danish universities - and a few research institutions - are involved in the research activities funded by the Ministry of Science, Technology and Innovation. University of Aarhus with the Danish School of Education and the Danish Clearinghouse for educational R&D holds a central position. Moreover, the 8 university colleges in partnership with regional knowledge centres, local authorities and schools are responsible for developmental activities financed by the Ministry of Education.

The current institutional framework is a result of two major institutional reforms, which were undertaken in 2000 and 2006 respectively.

First the article gives a brief description of the institutional reform undertaken in 2000. The reform was followed by a number of changes in the university sector and the OECD review of the Danish educational R&D in 2004. Then the article focuses on two processes that affected the Danish R&D in 2006: The governments Globalisation Strategy and the institutional reform in the university sector.

Institutional changes

In 2000, one of the central elements in the reform was the establishment of the Danish University of Education (DPU) by a merger of the Royal Danish School of Educational Studies, the Danish School of Advanced Teaching and the Danish Educational Institute. The idea was that the single institution, over a number of years, would become a leading international university in research and training in educational theory and practice, learning and competence development. Another of DPUs duties was to disseminate research-based knowledge to a wider public outside the academic world, including teachers at all levels of the educational system.

An independent research and development centre called Learning Lab Denmark, LLD, was merged into DPU as well. The purpose of LLD is to conduct research and development that transcends the strict borders of different academic disciplines, e.g. anthropology, political science, psychology, organisational and cultural theory and educational theory and practice, and to transcend the strict borders of educational and business sectors.

With the establishment of DPU the largest environment for academic research in didactics and education in Denmark was created. In a European context the university was also unique. In 2006, the university employed approximately 200 researchers in permanent positions and was hereby employing more than a third of the researchers (calculated as full year's work for one person) conducting educational studies in Denmark. Approximately 5.000 students were enrolled in study programmes, making the university the largest in Europe within the field. DPU was a leading partner in the ASEA-project: The ASIA-EUROPE Education and Research Hub for Lifelong Learning.

Another central element of the reform launched in 2000 was the merger of a wide range of institutions in the college sector into new Centres of Higher Education – university colleges. The university colleges offer bachelor degrees for teachers at the primary school level, nurses and other professions in the health sector as well as degrees in engineering and business. In the legislation of the university colleges, it was envisaged that they were "research affiliated" and this was ensured through cooperation agreements between university colleges and relevant research institutions in Denmark (as well as other countries) on user inspired research projects.

The DPU got a specific duty to support university colleges to be included in R&D projects. The purpose of the "research affiliation" of the university colleges was to ensure that the university colleges had ready and on-going access to recent research-based knowledge, to develop qualifications of CVU teaching staff and to build bridges between "scientific research" and "applied research and development".

Since 2004 a number of regional knowledge centres have been established. The purpose of the centres is to collect, process and transmit knowledge established by institutions offering short-cycle and medium-cycle study programmes to researchers, teachers, students, field workers, parents etc. The regional centres apply and distribute user-inspired knowledge developed by universities – in corporation with university colleges. A large share of the knowledge centres is involving university colleges with special competences of didactics. This is for example the case in relation to The Centre for Reading Research, The Centre for Curriculum Education and Development, The Centre for Adult Learning and Education. The Ministry of Education is funding the centres. The annual budget in 2007 is nearly 50 million DKK.

OECD review of educational Research and Development

In the university sector a number of significant institutional changes have been made during the last five years. In 2003 the government implemented a major reform of Danish

universities. The two central elements were the introduction of external board and employed – instead of elected – rector, deans and head of departments.

Part of the reform process was to request OECD to carry out a review of the Danish university sector and the sector for university colleges. The only OECD-recommendation on R&D was to strengthen educational R&D.

For this reason the Minister of Science, Technology and Innovation requested the OECD to examine national educational R&D. Prior to the OECD-review Denmark carried out a quantitative impact study of the national educational R&D which – together with a similar mapping made in 1999 – gave en overview of the development in educational R&D regarding research subjects, methods etc.

The OECD-review took place in 2004. The purpose of the review was to assess the extend to which it serves its function of creating, collating, distributing and applying the knowledge on which practitioners and policy makers can draw. Thus the aim of the examination was broader and different from a traditional educational R&D review that focuses on the quality of the research. The examination focused in particular on interactions between producers and users of educational R&D.

OECD concluded that there was no explicit national strategy for educational R&D. The R&D activities were organized in a large number of small scale projects largely determined by the individual interests of the researchers. Coordination and dialogue between the various actors – especially between researchers at the universities on one hand, and trainee teachers at the university colleges on the other - were modest. In general, there appeared to be a lack of basic research, a low capacity to apply a range of research methods (in particular quantitative methods) and few mechanisms for accumulating and sustaining the knowledge.

Another conclusion was that Denmark spend a large part of the national budget on education compared to other OECD countries, but used only a relatively minor portion of the resources for research and development within the educational field. The national spending on educational R&D amounted to 0.15 per cent of annual total public expenditure on education in Denmark, while the OECD average spending on educational R&D was about 0.27 percent of total expenditure on education.

Based on these conclusions and on experiences from US and UK OECD suggested the establishment of following new institutions:

- A National Education Research Forum, which could hold regular meetings (annually, or two or three times a year) for researchers, teachers and policymakers to share ideas on education issues and needs.
- A clearinghouse on education, which could be a means to offer regular reporting on significant educational research gained nationally as well as internationally
- An Educational observatory in Denmark

In the wake of the OECD report the Ministry of Education and the Ministry of Science, Technology and Innovation initiated a process of round table discussions to clarify the organisation of the new institutions.

At that point there seemed to be a general acceptance of the need to build new institutions at the national level in order to develop a national research strategy and strengthen coordination between actors. However, at the same time there was widespread disagreement on strategies, priorities and needs.

A Strategy for Denmark in the Global Economy

In April 2006 the Danish Government launched a Globalisation Strategy, "Strategy for Denmark in the Global Economy" [5]. The strategy was the outcome of a process featuring 14 meetings in the Globalisation Council – a council set up by the government comprising 26 key decision makers and chaired by the Prime Minister.

The strategy contained 350 specific initiatives, launching a comprehensive reform within the field of education, research and innovation affecting actors from primary schools to higher education, from public research institutions to private businesses.

The strategy focused on improving the efficiency of the public spending on education and research, in particular by allocating more public funds in open competition, and by increasing competition and internationalisation in the Danish economy as a whole. Competitiveness was perceived as a means to ensure prosperity and cohesion in the society.

A Globalisation Fund will finance the initiatives. DKK 10 billions will be set aside for future investments until 2012. The pledge will get gradually phased in, with another DKK 2 billion allocated per year until 2010 and DKK 1 billion allocated in 2011 and 2012.

University mergers

One initiative in the Globalisation Strategy was the proposal of a major institutional reform directed towards the university sector. The ambition was to create world-class universities by the merger of 13 government research institutions into the universities. At the same time the government instituted a merger process for the existing 12 universities and as of 2007 Denmark has 3 large universities, 4 medium-sized universities and one small university.

Danish University of Education has merged with University of Aarhus as a school of education with its own Advisory Board and Dean. The University of Aarhus is a modern university with a unique research profile combining basic research, user-driven research and research services for public authorities. The University of Aarhus is the broadest Danish University present in all regions of Denmark, including Copenhagen. The Education School will be part of a STAR Alliance with 10 universities in Asia, US, Canada and Australia.

Another institutional change made in 2006 was the merger of the University Colleges into 8 larger university colleges with coherent governance structure similar to the one introduced to universities in 2003.

The new institutional framework is expected to contribute to the building of a stronger relation between the university colleges and the university sector on user-inspired research projects.

A strong R&D base

In December 2006, the Danish government decided to set up a national committee on educational R&D. The committee was asked to describe models that enable close cooperation between users and producers of educational R&D.

The committee presented its report to the government in June 2007[6]. The report contains 26 recommendations on quality of educational R&D directed towards universities, university colleges and the government. The committee emphasizes that university R&D research is of high quality and focused on the needs of the educational sector.

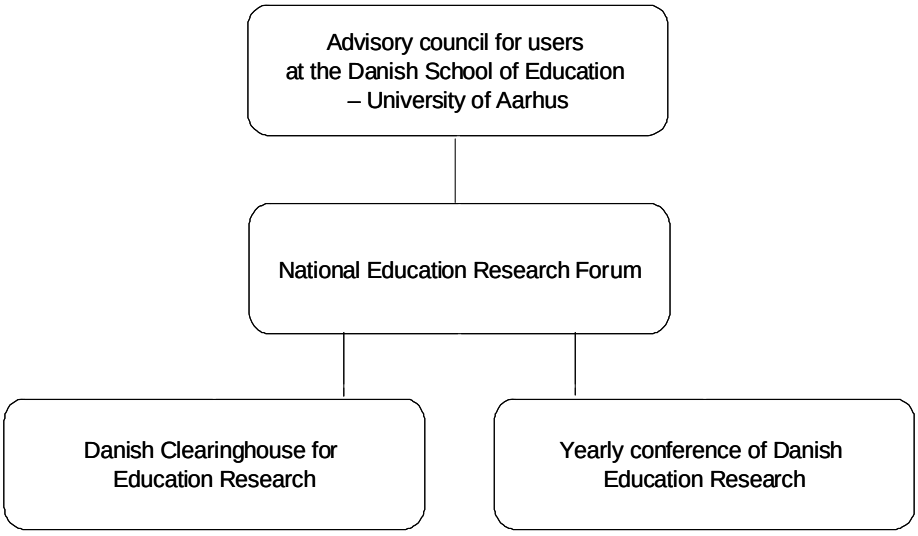
---

5 http://www.globalisering.dk/page.dsp?area=52

6 http://www.ubst.dk/page.php?pagegroupid=11&id=37

A central result of the work of the committee is that the ideas of a National Education Research Forum and clearinghouse are further developed. Today there is therefore a much stronger and broader consensus for the National Forum and Clearinghouse than in 2004 when the ideas were first presented as a follow-up to the OCED-review. In effect, the National Forum and the Danish Clearinghouse will build on a concentrated and stronger R&D-foundation for educational R&D in Denmark due to institutional research.

Suggested organisation of the National Research Forum and the Danish Clearinghouse for Education Research



The Ministry of Education and the Ministry of Science, Technology and Innovation are currently reviewing the committee's recommendations with a view to initiate new government initiatives for a stronger R&D base.

It has taken more than 5 years to build an institutional set-up that expectantly will allow Denmark to create world-class educational R&D and to apply and disseminate knowledge in learning to all sectors in the Danish educational

The Ministry of Science, Technology and Innovation and The Ministry of Education are looking forward to see the first national strategy from the National Educational Forum some time during the coming year.

One crucial aspect of that strategy has to be: How to document the strong Danish learning culture. For that reason we in Denmark would like to work towards a broad understanding of concept of evidence in education policy.

National Experience Report: The Netherlands

Hans Stegeman, Senior Policy Advisor for International Policy, Ministry of Education, Culture and Science

Introduction

The Netherlands are grateful for this opportunity to present to you a view on "Knowledge for Action".

The view which I will present today is partly a "national experience report" on how the Netherlands are trying to strengthen strategies for an evidence-based education policy.

Partly, too, it is a brief exposition on how the Netherlands, together with other countries and with international organisations, wish to contribute to the international effort for strengthening such strategies.

Specifically, I thank the organizers of this symposium and the German Presidency for creating this fine opportunity to exchange thoughts on matters that should occupy all of us.

Also I should mention with gratitude the group of interested countries who sent experts to two small seminars which the Netherlands organized in The Hague, in June and December last year, to further discuss practical approaches to evidence-based policy research. Colleagues from Austria, Denmark, Finland, Germany and the United Kingdom were present, as well as experts from the OECD and the European Commission. Specifically, I wish to mention my Danish colleagues, with whom I have had the pleasure of preparing a pilot project on which I will say something more near the end of my presentation.

The case of the Netherlands, I: advice of the "Education Council"

In the Netherlands, our debate on strategies for an evidence-based education policy got underway in 2006 through an advice of the "Education Council". This council is the authoritative expert body which advises our government on education matters.

The "Education Council" concluded in its 2006 advice that not enough effort was being made to gather proof for the effectiveness of educational methods – let alone to effectively use such proof.

The Council recommended, among other things:

- To put evidence-based strategies explicitly on the national policy agenda;
- To make relevant educational knowledge more accessible to researchers, policy makers, teachers;
- To stimulate evidence-based processes in schools themselves;
- To bring evidence-based components into teacher training.

In a policy reaction, last December, the Minister took up important elements from the advice.

- The policy reaction mentioned the founding of a "Knowledge Chamber" ("Kenniskamer"), as interface between policy, research and – ideally – practice. I will get back to this Chamber later on.
- Also, the minister mentioned the Dutch participation in the OECD-project "Schooling for tomorrow". A major aim for our participation could be summarized as "sharing knowledge for innovation". Two questions are considered central here: how to

improve the interaction between research and practice – and how to strengthen the exchange of knowledge between schools.

- The policy reaction also mentioned the need to guarantee transparency, monitoring and evaluation when new teaching methods are introduced.

A specific project mentioned in the policy reaction is the in-depth pilot project on "training schools" and "academic schools".

In the Netherlands a number of "Training schools" or "academic schools" offer new chances for linking teacher training, school development, educational innovation and research. Essentially, "training schools" are normal schools – but with a special teacher training function and connecting educational practice to an academic and training approach.

In these schools the objective is a synergy between the teaching of pupils, teacher training "on the spot" and research – the three mutually strengthening each other. This offers a perspective for teachers to become more involved in gathering and using "evidence" on how education functions.

The case of the Netherlands, II: participating in the international debate

Parallel to the national development of the theme, the Netherlands participated in the international project that had already been started by the Centre for Education Research and Innovation (the OECD's "CERI") and which led to the successful London conference of July 2006.

Besides, next to our participation in the "evidence-based policy-project" of the OECD's "CERI", the Netherlands organised the two seminars in The Hague which I mentioned at the beginning.

Our aim was to consider whether we could join efforts with other countries to strengthen the approach to evidence-based strategies. We were glad to find that the European Commission is very committed to the theme – as is the German Presidency.

We welcome the possibilities we see to combine our national effort with the international efforts that are under way.

An eco-system of knowledge

In the Netherlands, our thinking on strengthening evidence-based education strategies is linked to a conceptual framework which we call the "eco-system of knowledge".

Although this conceptual framework reflects the Netherlands' national experience on policy-making (evidence-based or not) it is not at all a Dutch invention. In fact the framework will, I imagine, seem familiar to many of you. It corresponds closely to concepts as they are coined in the OECD debate on evidence-based policy making.

This ecosystem essentially consists of the *triangle* of

- Policy,
- Research and
- Practice.

In fact, these three subsystems cover most of the groups and institutions concerned with educational policy making:

- Policy: policy-makers, chief-scientists, analysts, politicians;
- Research: researchers, universities, companies, development centres, -knowledge centres;
- Practice: teachers, schools, unions and increasingly pupils and parents.

Ideally, the three sides of the triangle would coincide with three "streams of interaction" between the subsystems.

- There would be the interaction between policy and practice. But in fact, we often see rather a one-sided top-down approach instead of interaction;
- We would also hope for the interaction between policy and research. This interaction exists in most cases – but is it adequate?
- And there should be the interaction between research and practice. We often see a one-sided instead of an interactive relationship. How often do researchers (or politicians) use evidence generated by teachers? And *vice versa*: Do teachers have access to research results; still stronger: are they at all interested? The concept of "academic schools" which I mentioned earlier might be useful here.

The case of the Netherlands, III: developing new instruments

In the Netherlands, growing awareness that new dynamics are needed within this ecosystem (following on the 2006 advice of the Education council) has led to the developing of new instruments. One is a new unit within the Ministry of Education Culture and Science, which is currently being staffed. The other, much more important, is a new form of "interface" between the Ministry and the other players in the ecosystem; it is called a "Knowledge Chamber" (*Kenniskamer*). Let me briefly say something more about both.

The *Knowledge Chamber* is partially an expedient political response to recently expressed doubts concerning educational policy-making. In the Dutch context it also represents a time-honoured method of tackling problems - bringing together government and stakeholders in a structuralized give-and-take of views, information and knowledge. It has met twice, so far.

Policy questions are put before the Knowledge Chamber. The Chamber creates conditions for generating, offering and obtaining knowledge to deal with these policy questions on a basis of evidence.

As it will gain momentum, the Knowledge Chamber will meet in two forms:

- The regular meetings will be attended by an inner circle composed of the government's main advisory bodies on education and research as well as the inspectorate.
- For the theme-oriented special meetings also an outer circle will be invited, of varying composition.

The Knowledge Chamber aims to become the nucleus of a network of knowledge workers and policy-makers. The Knowledge Chamber will generate *validated* knowledge which may be used by policy makers.

Yet in its current phase, after having met twice, the Knowledge Chamber mainly represents only one side of the triangle: the side connecting research and policy-making. Educational practice seems, as yet, to be under-represented.

The *new unit* within the Ministry is the "Knowledge Directorate". It is currently under construction. The new directorate will:

- Stimulate a sustainable exchange of knowledge between the partners within the knowledge ecosystem;
- Underpin and substantiate policy plans with outcomes of scientific research and good practice;
- Link economy-driven demand for innovation to developments in the field of education, science and culture.

This new directorate, though still under construction, is already vigorously at work.

Perspectives for further debate

Quite recently, a seminar within the Dutch Ministry helped set the agenda for the new Knowledge directorate. Some very interesting conclusions of the discussion were the following:

- Admitting new players – *e.g.* parents and pupils - into the ecosystem can lead to radical differences in concepts and demands for policy-making;
- State-of-the-art experiments are not feasible within the living context of education. Therefore, other forms of research are needed to open up the "black box" of education and find the wished-for evidence. International cooperation widens the possibility of finding evidence from practices elsewhere ("natural experiments");
- Grassroots experience can help generate validated knowledge. Teachers in their classrooms may be seen as researchers, too. They gather evidence;
- And the instrument of meta-analysis of existing research outcomes will help informed decision-making and strengthen the evidence-based foundation of policy-making.

An international pilot project proposal; concluding remarks

Ladies and Gentlemen,

Earlier, I mentioned the pilot project which Denmark and the Netherlands propose starting up, together with the five other countries which sent experts to the 2006 seminars in The Hague.

I would like to make some brief remarks on the proposed pilot.

At the The Hague seminars, we were gratified by the interest from the European Commission, which since has proposed to help us organize a "Peer Learning Activity" (PLA) on evidence-based education policy, in the Netherlands – an offer which we accept eagerly: we aim at organizing the PLA in May.

We were enthusiastically supported in all this by our colleagues from Denmark who, after our latest pilot-seminar in The Hague, December last, agreed to set the ball rolling together with us: we intend, this spring, to perform the first steps in our pilot project.

The pilot project aims to perform during the year 2007, a number of bilateral case studies in order to contribute to a European knowledge-base and to stimulate the use of validated knowledge in the developing of educational strategies in the EU.

Each case study will involve a target country and an examining country. We expect that – at least – the countries with which we held the two seminars in 2006 will participate in the pilot, each performing a case study as well as undergoing one.

As mentioned before, Denmark and the Netherlands intend to set the ball rolling. The Netherlands will perform a case study on the topic proposed by our Danish colleagues. The topic proposed is the introduction of national test schemes in the primary and lower secondary schools in Denmark and the establishment of the "National Authority for Quality Assurance and Evaluation of Primary and Lower Secondary Schools". The knowledge base and the source for evidence in the process will be PISA, The Review of Quality and Equity in Danish Schools performed by the OECD and the work of thematic working groups on the recommendations in the review.

The topic proposed by the Danish colleagues is of great present interest to the Netherlands, as we are currently experiencing a debate on the merits and risks of testing primary education pupils.

The Netherlands and Denmark will, of course, inform you of the outcome of this process and will also stimulate that the other interested countries – the "The Hague Group" – will perform a similar case study exercise.

This brings me to my concluding remarks.

A national debate on strengthening the evidence-base in education and education policy has started in the Netherlands in 2006. The advisory "Education Council" played an important role. Since its start the debate has widened and more players are involved. This is a positive development.

The Netherlands' debate is part of a wider international debate. Both the European Commission and the OECD, but also experts from individual countries, are providing input for this international debate. We feel very strongly that this international context is crucial to develop an evidence-based education policy; if only because the international dimension widens the scope of evidence-gathering.

An eco-system of knowledge, a triangle composed of practice, policy and research, may offer a conceptual framework to work towards stronger evidence-based strategies. With the help of this framework, we have concluded in the Netherlands that the dynamics of interaction between our three subsystems are still very incomplete.

But we are working on it. The Knowledge Chamber and the new Knowledge directorate are proof of this. So is our wish to learn from international partners and to participate in international projects. But there is still much to do – both in the Netherlands and elsewhere.

Thank you.

Intervention studies: Experimental approaches and alternatives

Rolf Steyer, Ulf Kroehne and Christiane Fiege

Introduction

For more than a century there have been examples in the statistical literature showing that comparing means or comparing probabilities (e.g., of success of an intervention) between a group exposed to an intervention and a comparison group (unexposed or exposed to a different intervention) does not necessarily answer our substantive questions: 'Which intervention is better overall?' or 'which intervention is better for which kind of persons?' Mean differences, differences in probabilities or any other comparisons between means or probabilities are usually not the causal intervention effects we are looking for (see, e.g., Yule, 1903; Simpson, 1951). They are just 'effects at first sight' or "prima facie effects" (cf. Holland, 1986).

The Simpson paradox

In this section, we will treat an example in which the prima facie effects invert if we switch from comparing relative frequencies (between the intervention group and the comparison group) in the total sample to comparing the corresponding relative frequencies in subsamples such as 'male' and 'female'. This kind of phenomenon is called the *Simpson paradox* (Simpson, 1951).

Table 1*: Frequencies in the total sample*

	Intervention		
Success	No (X=0)	Yes (X=1)	
No (Y=0)	240	232	472
Yes (Y=1)	360	168	528
	600	400	1000

Prima facie effect in the total sample. Table 1 presents data on 1000 subjects, 400 of which participated in an intervention ($X = 1$) and 600 of which did not participate ($X = 0$). As shown in this table, 168 of the participants were successful ($Y = 1$), the other 232 were not ($Y = 0$), whereas 360 of the non-participants were successful and the other 240 were not. Comparing the proportions of success between those which were exposed to the intervention, 168/(168 + 232) = .42, and those which were not, 360/(360 + 240) = .60 (see Fig. 1), may lead to the conclusion that *the intervention is harmful*.

Prima facie effects in the subsamples. This conclusion will be completely different if we look at the same data separately for males and females. Table 2 shows that 16 of the treated 16 + 4 = 20 males (or 80%) were successful whereas 336 of the nontreated 336 + 144 = 480 males (or 70%) were successful. Considering the difference 80%−70% = 10% may lead us to conclude that *the intervention is bene• cial for males* (see also Figure 2).

What about the intervention effects in the female sample? According to our intuition, there should be a big negative effect of the intervention for the females in order to

compensate for the positive effect in the male sample so that there still can be a negative effect in the total sample. However, our intuition fails (in this example). Table 2 shows that 152 of the treated 380 females – this is 40% – were successful whereas only 24 of the non-treated 120 females – this is 20% – were successful. Considering the difference 40% – 20% = 20% may lead us to conclude that *the intervention is also beneficial for females* (see again Figure 2).

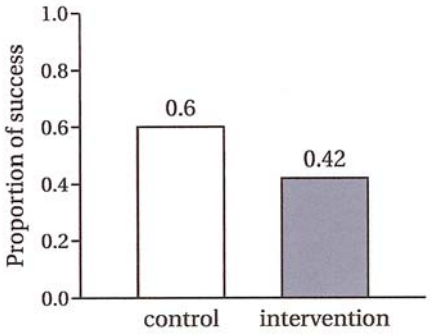


Figure 1*: Proportions of success given intervention in the total sample Proportion of success*

Table 2*: Frequencies in the subsamples of males and females*

Males (Z=0)

Success	Intervention: No (X=0)	Intervention: Yes (X=1)	
No (Y=0)	144	4	148
Yes (Y=1)	336	16	352
	480	20	500

Females (Z=1)

Success	Intervention: No (X=0)	Intervention: Yes (X=1)	
No (Y=0)	96	228	324
Yes (Y=1)	24	152	176
	120	380	500

Hence, the intervention seems to be *beneficial for both, males and females.* This seems to contradict our finding in the total sample consisting of these males and females. There, the *intervention seems to be harmful*. In contrast to our intuition, the prima facie effect in the total sample is *no*t the average of the prima facie effects in the two subsamples of males and females. Instead the effect in the total group is inverted if we look at the effects in the subgroups of males and females of which the total group consists.

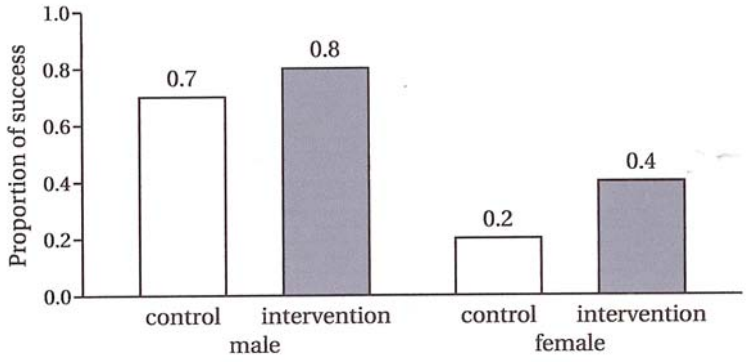


Figure 2*: Proportions of success given intervention within gender*

Implications of the paradox

One might think that the contradictory conclusions described above might be due to a peculiar, small sample and would not occur in the population. However, exactly the same problem pertains if we increase the sample size by any factor – as long as each of the sample sizes in the cells of Table 2 is increased by the same factor. Hence, instead of the 4 unsuccessful males that are exposed to the intervention we could have 4000 and instead of the 16 successful males that are exposed to the intervention we could have 16000 etc. The proportions presented in Figures 1 and 2 will not change this way and our conclusions remain the same: Considering the total group, there is a negative effect of the intervention, looking at each of the two subgroups yields a positive effect. Hence, this phenomenon is not a problem of statistical inference; it also occurs at the population level and it should be noted that exactly the same problem also occurs if, instead of a dichotomous outcome variable (in our example: success), we consider a quantitative outcome variable. In this case we would compare means between groups instead of probabilities.

Since the conclusions drawn from the comparisons in the total population and within the gender subsamples are contradictory, which of these comparisons should we trust? Is the intervention harmful – as the comparison in the total sample suggests? Or is it beneficial as suggested by the comparisons within the gender subsamples? Which of these comparisons are meaningful for the evaluation of the causal effects of the intervention?

These questions cannot be answered without being more explicit in our goals. What are we looking for? If it really is the evaluation of the effects of the intervention, we will have to define the concept of a *causal effect*. Just comparing probability or mean differences in an outcome variable between intervention conditions does not tell whether the intervention is good or bad for the subjects. But this is our substantive interest: finding out if the intervention is harmful or if it is beneficial!

Obviously, the inversion of effects between the total sample and subsamples may also occur if we split up these subsamples again. For example, in subsamples such as 'young' and 'old' of the male sample we might have a new inversion of effects. Since this phenomenon also occurs at the population level, neither the probability (or mean) difference in the total population, nor in subpopulations can be used to define a causal effect, and this conclusion can be generalized to all kinds of comparisons of means, probabilities and any other parameters describing the distribution of an outcome variable.

How to escape from this dilemma? Obviously, an inversion of effects is not possible any more if we built our theory on the smallest possible subpopulations: the observational units (in the social and behavioral sciences, usually the subjects or individuals) which may be exposed to an intervention or not and for which we assess some outcome measure. Traditional statistical models may include subpopulations, but they do not go beyond, i.e.,

they do not incorporate the observational units and this is why these models are not informative about causal intervention effects *per se*. However, under certain conditions, they *can* be informative about causal intervention effects. But which are these conditions? How can we create them in experimental and quasi-experimental intervention studies? And which strategies of data analysis allow to estimate these effects? Obviously, conclusive answers to these questions can be hoped for only within a theory of causal effects.

Of course, raising these questions and tempting answers is not new. Many readers will be familiar with the Campbellian tradition of experiments and quasi-experiments (see, e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). Immense knowledge and wisdom about experiments and quasiexperiments has been collected within this tradition. Nevertheless, a more formal approach has been developed in the last century supplementing the Campbellian theory and terminology in important aspects: the theory of individual and average causal effects in the Neyman-Rubin tradition (see, e.g., Neyman, 1923/1900; Rubin, 1974). Many papers and books indicate the growing influence of this theory (see e.g., Rubin, 2006; Winship & Morgan, 1999) and formidable efforts have been made to integrate it into the Campbellian tradition (West, Biesanz, & Pitts, 2000).

In the remaining sections of this paper we give a short outline of this theory and discuss its consequences for experimental design and strategies of data analysis. For a more complete presentation we refer the reader to Steyer, Partchev, Kröhne, and Nagengast (2007).

Theory of individual and average causal effects

The basic concepts of the theory of individual and average causal effects will be explained by an example. We will first look at the basic concepts of the theory itself, and then study how they are related to the parameters, such as means, that are usually estimated in empirical studies. For simplicity, we will restrict our presentation to the case of two intervention conditions, one of which might be a control.

Basic concepts

Table 3 displays a population of six observational units (persons). The first four units are males $(Z = m)$, the other two are females $(Z = f)$. In the random experiment to which the theory refers, one of these six units is sampled with a probability of 1/6 (see second column) and is then either exposed to the intervention [with probability $P(X = 1 \mid U = u)$; see the last column of the fundamental parameters] – or not [with probability $1 - P(X = 1 \mid U = u)$]. Finally, the outcome for the person is observed, which is a value of the outcome variable *Y*.

True outcomes. It is presumed that there is a unit-specific distribution for both, the intervention and the no-intervention conditions. The unit-specific distribution of a person u in the intervention condition, for example, consists of the probabilities of all values of the outcome variable Y which could occur if this person would be exposed to the intervention. Furthermore, we presuppose that these unit-specific distributions have expectations: the *true outcome*

$$\tau_0(u) \equiv E(Y \mid X = 0, U = u) \qquad (1)$$

of unit *u under no-intervention* and the *true outcome*

$$\tau_1(u) \equiv E(Y \mid X = 1, U = u) \qquad (2)$$

of the same unit *under the intervention*. In the example presented in Table 3, the true outcome of person $u1$ in the no-intervention condition is 68 and its true outcome under the intervention is 81. Note that $E(Y \mid X = j, U = u) = P(Y = 1 \mid X = j, U = u)$, if the outcome variable Y is dichotomous with values 0 and 1, such as in our introductory example.

Table 3*: Example illustrating the basic concepts*

	Fundamental parameters					Derived parameters		
Unit u	Sampling probability $P(U=u)$	Covariate gender Z	True outcome under control $\tau_0(u) = E(Y \mid X=0, U=u)$	True outcome under intervention $\tau_1(u) = E(Y \mid X=1, U=u)$	Individual intervention probability $P(X=1 \mid U=u)$	Individual causal effect $\delta_{10}) = \tau_1(u) - \tau_0(u)$	$P(U=u \mid X=0)$	$P(U=u \mid X=1)$
u_1	1/6	m	68	81	3/4	13	1/10	3/14
u_2	1/6	m	78	86	3/4	8	1/10	3/14
u_3	1/6	m	88	100	3/4	12	1/10	3/14
u_4	1/6	m	98	103	3/4	5	1/10	3/14
u_5	1/6	f	106	114	1/4	8	3/10	1/14
u_6	1/6	f	116	130	1/4	14	3/10	1/14
Expectations			92.33	102.33	7/12	10.00		
Average causal effect (*ACE*)						10.00		
Prima facie effect (*PFE*)						−3.086		
					male	female		
Conditional *ACE*s					9.50	11.00		
Conditional *PFE*s					9.50	11.00		

Individual causal effects. The difference

$$\delta_{10}(u) \equiv \tau_1(u) - \tau_0(u) \qquad (3)$$

between these unit-specific expectations defines the *individual causal effect* of the intervention as compared to the no-intervention condition (see the first column of the derived parameters in Table 3). Hence, the individual causal effect of person u_1 is 13.

Let us emphasize that knowing an individual causal effect is of crucial interest in many situations. A teacher or a therapist, for instance, would like to know the true outcome of an intervention as compared to no intervention (or to an alternative intervention) for the individual person he is just working with. The same is true for physicians, social workers and everybody working with individuals. In practice, however, the unit can often be assigned only to one single experimental condition. For instance, a student cannot be simultaneously taught mathematics by a new method, represented by $X = 1$, and by a traditional teaching method, represented by $X = 0$. Therefore, it is often impossible to estimate the individual causal effect $\delta_{10}(u)$ for a given person u. Hence, in these cases we can either estimate the true outcome $\tau_1(u)$or $\tau_0(u)$, but neither both nor their difference. This has been called the *fundamental problem of causal inference* by Holland (1986), and this is also the most important reason why the average causal effects are of interest.

Average causal effects. Taking the expectation of the individual causal effects then defines the *average causal effect* of the intervention as compared to the no-intervention condition, i.e.:

$$ACE_{10} \equiv E(\delta_{10}). \tag{4}$$

Since $E(\delta_{10}) = E(\tau_1 - \tau_0) = E(\tau_1) - E(\tau_0)$, the average causal effect can also be computed by

$$ACE_{10} = E(\tau_1) - E(\tau_0), \tag{5}$$

i.e., by the difference between the two expectations of the true-outcome variables $E(\tau_1)$and $E(\tau_0)$. In our example,

$$E(\tau_0) = 68 \cdot \frac{1}{6} + \ldots + 116 \cdot \frac{1}{6} \approx 92.33 \tag{6}$$

and

$$E(\tau_1) = 81 \cdot \frac{1}{6} + \ldots + 130 \cdot \frac{1}{6} \approx 102.33 \tag{7}$$

respectively, and their difference is 10, which is also the expectation

$$E(\delta_{10}) = 13 \cdot \frac{1}{6} + \ldots + 14 \cdot \frac{1}{6} = 10 \tag{8}$$

of $\delta_{10} \equiv \tau_1 - \tau_0$ (see the row 'Expectations' in Table 3).

Although the ACE is the expectation of the corresponding individual causal effects, we will show that there are ways to identify the ACE *without knowing the individual causal effects* and without knowing the individual true outcomes in both intervention conditions, provided that we can make the necessary assumptions based on a well-designed study or on theoretical considerations and a careful and informed choice of the covariates.

Average causal effects given a value of the covariate. Sometimes we are not only interested in the average causal effect in the total population, but also in the average causal effect of the males or of those with a high pretest score. Taking the conditional expectation of the individual causal effects defines the *average causal effect* of the intervention as compared to the no-intervention condition *given a value of the covariate*, i.e.:

$$ACE_{10;\,Z}(z) \equiv E(\delta_{10} \mid Z = z) \tag{9}$$

This effect can also be computed by

$$ACE_{10;Z}(z) = E(\tau_1 \mid Z = z) - E(\tau_0 \mid Z = z), \tag{10}$$

i.e., by the difference between the two conditional expectations of the true-outcome variables $E(\tau_0 \mid Z = z)$ and $E(\tau_1 \mid Z = z)$.

In our example, the conditional expectations of the true outcomes for males are

$$E(\tau_0) \mid Z = m) = 68 \cdot \frac{1}{4} + \ldots + 98 \cdot \frac{1}{4} = 83 \tag{11}$$

and

$$E(\tau_1) \mid Z = m) = 81 \cdot \frac{1}{4} + \ldots + 103 \cdot \frac{1}{4} = 92.50 \qquad (12)$$

and their difference is 9.50. For the females these conditional expectations are

$$E(\tau_0) \mid Z = f) = 106 \cdot \frac{1}{2} + 116 \cdot \frac{1}{2} = 111 \qquad (13)$$

$$E(\tau_1) \mid Z = f) = 114 \cdot \frac{1}{2} + 130 \cdot \frac{1}{2} = 122 \qquad (14)$$

and their difference is 11.00 (see Table 3).

Comparing theoretical concepts to estimable parameters

Which is the relationship between the theoretical concepts and estimable parameters? How are the expectations $E(\tau_0)$ and $E(\tau_1)$ of the true outcomes in the two intervention conditions related to the corresponding conditional expectations $E(Y \mid X = 0)$ and $E(Y \mid X = 1)$ of the outcomes that are estimated by the sample means in empirical applications? Knowing the answer will also show how the average causal effects $ACE_{10} = E(\tau_1) - E(\tau_0)$ are related to the prima facie effects $PFE_{10} \equiv E(Y \mid X = 1) - E(Y \mid X = 0)$. Similarly, knowing how the conditional expectations $E(\tau_0 \mid Z = z)$ and $E(\tau_1 \mid Z = z)$ of the true outcomes are related to the corresponding conditional expectations $E(Y \mid X = 0, Z = z)$ and $E(Y \mid X = 1, Z = z)$ will also show how the average causal effects $ACE_{10;\,Z}(z) = E(\tau_1 \mid Z = z) - E(\tau_0 \mid Z = z)$ are related to the prima facie effects $PFE_{10;\,Z}(z) \equiv E(Y \mid X = 1, Z = z) - E(Y \mid X = 0, Z = z)$.

Expected true outcomes of an intervention. The *expected true outcome of intervention j* is:

$$E(\tau_j) = \sum_u \tau_j(u) \cdot P(U = u). \qquad (15)$$

Hence, $E(\tau_j)$ is the sum of the true outcomes $\tau_j(u)$ of unit u in intervention j weighted by their *unconditional* probabilities $P(U = u)$. It is this formula that we already used in Equations (6) and (7) yielding $E(\tau_0) = 92.33$ and $E(\tau_1) = 102.33$ Hence, the average causal effect is: $ACE_{10} = 102.33 - 92.33 = 10.00$.

Conditional expectations of outcomes given an intervention. Which are the conditional expectations $E(Y \mid X = 1)$ and $E(Y \mid X = 0)$ of the outcomes in the intervention and no-intervention conditions that would be estimated by the means $\overline{y_1}$ and $\overline{y_0}$ of the outcomes in a sample and that would be tested to be zero in an ordinary *t*-test for independent groups?

If the outcome variable Y is discrete, the expectations of Y given intervention j can be defined by: $E(Y \mid X = j) \equiv \sum_y y \cdot P(Y = y \mid X = j)$. However, the conditional expectations $E(Y \mid X = j)$ can always also be represented as the sum of the *true outcomes* $\tau_j(u)$ weighted by their *conditional probabilities* $P(U = u \mid X = j)$, i.e.:

$$E(Y \mid X = j) = \sum_u \tau_j(u) \cdot P(U = u \mid X = j). \qquad (16)$$

In the example presented in Table 3, we assume a probability of 3/4 that a male unit is exposed to the intervention, whereas the probability that a female unit is exposed to the intervention is 1/4 (see the last column of the fundamental parameters). If there is self-selection to the intervention conditions, such a difference between these individual probabilities might be due to the fact that the intervention is more attractive to males than to females. These differences in the individual probabilities $P(X = j \mid U = u)$ imply that also the probabilities $P(U=u \mid X = j)$ differ between the persons (see the last two columns of Table 3), because

$$P(U = u \mid X = j) = \frac{P(X = j \mid U = u) \cdot P(U = u)}{P(X = j)} \qquad (17)$$

Applying Equation (16) to the example for the intervention condition yields:

$$E(Y \mid X = 1) = \sum_u \tau_1(u) \cdot P(U) = u \mid X = 1)$$
$$= 81 \cdot \frac{3}{14} + \ldots + 103 \cdot \frac{3}{14} + 114 \cdot \frac{1}{14} + 130 \cdot \frac{1}{14} \approx 96.71$$

and $E(Y \mid X = 0) = 99.80$ for the no-intervention condition. Hence, the prima facie effect is: $PFE_{10} = 96.71 - 99.80 = -3.09$.

The source of bias of the prima facie effects. Comparing Equation (16) for the conditional expectation $E(Y \mid X = j)$ of the outcomes under intervention j to the *expected outcome* $E(\tau_j)$ *of intervention j* [see Eq. (15)] reveals that $E(\tau_j)$ is the sum of the true outcomes $\tau_j(u)$ weighted by their *unconditional probabilities* $P(U = u)$, whereas the corresponding conditional expectation $E(Y \mid X = j)$ is the sum of the true outcomes $\tau_j(u)$ weighted by their *conditional probabilities* $P(U = u \mid X = j)$ [see Eq. (16)]. In general

$$E(Y \mid X = j) = E(\tau_j) + bias_j, \qquad (18)$$

where $bias\, j \equiv E(Y \mid X = j) - E(\tau_j)$.

The difference between $E(Y \mid X = j)$and $E(\tau_j)$ is why the conditional expectations $E(Y \mid X = j)$ of the outcomes in the intervention conditions can be biased and misleading if used for causal inference. Whilst the expectations $E(\tau_j)$of the true outcomes in the intervention conditions do *not depend* on the individual intervention probabilities $P(X = j \mid U = u)$, the expectations $E(Y \mid X = j)$ of the outcomes in the intervention conditions *do depend* on these individual intervention probabilities [see Eqs. (17) and (16)]. Hence, if $P(X = j \mid U = u) = P(X = j)$ does not hold for each unit u, then $P(U = u \mid X = j) = P(U = u)$, will also not hold for each unit u, which means that we can neither expect $E(Y \mid X = j) = E(job)$ nor $PFE_{jk} = ACE_{jk}$,because $PFE_{jk} \equiv E(Y \mid X = j) - E(Y \mid X = k)$, whereas $ACE_{jk} = E(\tau_j) - E(\tau_k)$.

Identification of causal effects

Identification of the average causal effect ACE_{jk} is possible under two conditions:

(a) unbiasedness of the regression $E(Y \mid X)$ of Y on X , and (b) unbiasedness of the regression $E(Y \mid X, Z)$ of Y on X and Z, where Z denotes a covariate. The covariate Z can be uni-or multidimensional, manifest or latent. For the time being, this should be sufficient to know. Later on we will give more details on different kinds of covariates.

Identification **under unbiasedness of** $E(Y \mid X)$

The average causal effects can be identified via the prima facie effects *if the regression* $E(Y \mid X)$ *of* Y *on* X *is unbiased*, i.e., if

$$E(Y \mid X = j) = E(\tau_j), \text{ for each value } j \text{ of } X. \qquad (19)$$

If this equation holds for each intervention j , then the prima facie effect will also be *unbiased, i.e.*

$$PFE_{jk} = ACE_{jk} \qquad (20)$$

Therefore, if the regression $E(Y \mid X)$ is unbiased, we may use the prima facie effect PFE_{jk} for causal inference, i.e., we simply have to take the difference $E(Y \mid X = j) - E(Y \mid X = k)$ in order to compute the average causal effect ACE_{jk} .Hence, in a sample the mean difference $\overline{y_j} - \overline{y_k}$ is – under the usual linear model assumptions underlying the *t*-test – an unbiased estimate of the corresponding average causal effect, *provided that the regression* $E(Y \mid X)$ *is unbiased.*

Sufficient **conditions for unbiasedness of** $E(Y \mid X)$

Comparing Equation (15) to (16) immediately reveals a sufficient condition for unbiasedness of the conditional expectation $E(Y \mid X = j)$ of the outcomes given intervention j . If the probabilities $P(X = j \mid U = u)$ do not differ between observational units, i.e., if $P(X = j \mid U = u) = P(X = j)$, for each unit *u*, or in other words

$$P(X = j \mid U) = P(X = j), \qquad (21)$$

then this implies $P(U = u \mid X = j) = P(U = u)$ for each unit u [see Eq. (17)]. In this case, the bias of $E(Y \mid X = j)$ vanishes [see Eqs. (15) and (16)]. If Equation (21) holds *for each intervention j, then units and interventions are (stochastically) independent.* Hence, stochastic independence of units and interventions implies that the regression $E(Y \mid X)$ itself is unbiased. This is the most important sufficient condition for unbiasedness of the regression $E(Y \mid X)$. It is secured, for instance, by random assignment of the unit to one of the intervention conditions in the randomized experiment. If this equation holds, we can use Equations (19) and (20) to identify the expected true outcomes and the average causal effects, respectively.

There are at least four other sufficient conditions for unbiasedness of the regression, some of which are logically weaker than stochastic independence of units and interventions. One of these other conditions is stochastic independence of X and the vector of the true outcome variables.[7] Another one is that the true outcomes do not differ between units within each intervention condition j [see Steyer et al. (2007) for details and other sufficient conditions].

Identification **under unbiasedness of** $E(Y \mid X, Z)$

The average causal effects can also be identified *if the regression* $E(Y \mid X, Z)$ *of* Y *on* X *and* Z *is unbiased, i.e.*, if

$$E_{X = j}(Y \mid Z) = E(\tau_j \mid Z), \text{ for each value } j \text{ of } X, \qquad (22)$$

7 This condition corresponds to Rubin's 'strong ignorability' (see, e.g., Rosenbaum & Rubin, 1983).

where $E_X = _j (Y \mid Z)$ denotes the regression of Y on a covariate Z within intervention condition j . The reason why the *ACEs* can then be identified is that Equation (22) implies

$$E[E_X = j (Y \mid Z)] = E[E(\tau j \mid Z)] = E(\tau_j). \quad (23)$$

Estimates of $E[E_X = j (Y \mid Z)]$ are called *adjusted means* and we will refer to data analysis strategies based on this equation as 'adjustment procedures'.
According to this equation, we simply have to take the expectation $E[EX = j (Y \mid Z)]$ of the regression $EX = j (Y \mid Z)$ of Y on the covariate Z within an intervention condition j in order to compute the expected outcome $E(\tau_j)$, *provided that* $E(Y \mid X , Z)$ *is unbiased* [see Eq. (22)]. The difference between $E(\tau_j)$and $E(\tau_k)$ then yields the average causal effect of intervention j compared to a different (or no) intervention
k. Also note that if $E(Y \mid X , Z)$ is unbiased, the difference

$$ACE_{jk;\ Z}(z) = E_X = _j (Y \mid Z = z) - E_X = _k (Y \mid Z = z), \quad (24)$$

is the *average causal effect* of intervention j compared to intervention *k given the value z of the covariate Z* [see Eqs. (22) and (10)].

Sufficient **conditions for unbiasedness of** $E(Y \mid X , Z)$

A sufficient condition for unbiasedness of $E(Y \mid X , Z)$ is *conditional independence of units and treatments given the covariate Z* , i.e.,

$$P(X = j \mid Z) = P(X = j \mid Z ,U), \text{ for each value } j \text{ of } X . \quad (25)$$

If this equation holds, we can use Equation (23) to identify the expected true outcomes and the average causal effects by taking the differences between them. Furthermore, under conditional independence of units and treatments, we can also use Equation (24) to compute the average causal effect given a value z of the covariate. Condition (25) may be secured, for instance, by conditional random assignment of the unit to one of the intervention conditions in a conditionally randomized experiment.

As an example of *conditional* random assignment, suppose that there are three values of the covariate Z = *knowledge level*. The experimenter could then decide to treat student u with probability 5/6 if his or her knowledge is 'low', with probability 3/6 if knowledge is 'medium', and with probability 1/6 if knowledge is 'high'. In this case, the treatment probability is only a function of Z and will not depend on any other attribute of *u*. In this case Equation (25) will hold.

However, Equation (25) may also hold, if the experimenter does *no*t *contro*l the individual intervention probabilities, *provided that the covariate Z is adequately chosen*. Since Z can be multivariate, we 'only' have to choose the covariates $Z_1,\ldots, ZQ$ such that Equation (25) holds for $Z = (Z_1,\ldots, ZQ)$. Of course, this is a realistic assumption only if we have a valid theory about which covariate Z will fulfill Equation (25).

For a numerical example consider again Table 3. In this example, $P(X = 1 \mid Z) = P(X = 1 \mid Z ,U)$, because within the male subpopulation Z = m the individual probabilities $P(X = 1 \mid Z = m, U= u) = P(X = 1|U= u) = 3/4$ do not differ between each other. This implies that also the conditional probabilities $P(X = 0 \mid U= u) = 1-P(X = 1 \mid U= u)$ do not differ between units *u*. The same is true in the female subpopulation, where $P(X = 1|Z = f ,U= u) = P(X = 1|U= u) = 1/4$. This implies that the regression $E(Y \mid X , Z)$ is unbiased. Hence, in this example, the difference $EX = 1(Y |Z = m)-EX = 0(Y \mid Z = m) = 92.50 - 83.00 = 9.50$ is also the average causal effect of the intervention in the male subpopulation, and the difference $EX = 1(Y |Z = f)-EX = 0(Y |Z = f) = 122.00-111.00 = 11.00$ is also the average causal effect of the intervention in the female subpopulation.

Kinds of covariates

Obviously, covariates play an important role in identifying causal effects. Hence, observing those covariates which determine the probabilities that a unit is exposed to intervention j is of crucial importance in quasi-experiments. Without the correct covariates, the study is useless or even misleading when it comes to drawing causal inferences.

A covariate – in Table 3 'gender' – represents an attribute of the observational unit *before the intervention may exert its effects*. In this case the covariate is a function $f(U)$ of the observational-unit variable U. However, a covariate Z can also be an attribute of the unit measured with error. In this case $Z = f(U) + \varepsilon$, where ε is an error term. An example is a test score in an intelligence test. A covariate never represents an attribute or a fallible measure of such an attribute that might be affected by the treatment. Hence, *mediator variables*, which, by definition are affected by the treatment variable and affect themselves the outcome, *are never covariates*. As mentioned before, there are different kinds of covariates.

Manifest covariates. Covariates can be *manifest or observable*, i.e., they may be defined by a simple rule assigning the values of the covariate to the possible observations in the random experiment considered (see Section 2.1). Examples are the scoring rules of psychological tests, such as 'number of correct answers' in an ability test. A manifest covariate may also be *multidimensional*, i.e., it may consist of several unidimensional covariates which may be qualitative such as *gender*, or quantitative such as *ability* with test scores ranging between 0 and some maximum value. Both variables are unidimensional. Considered together, they form a two-dimensional manifest covariate.

Latent covariates. The covariate Z may also be latent and possibly multidimensional. Such a latent covariate can be defined in the frameworks of Classical Test Theory (CTT; Lord & Novick, 1968; Steyer, 2001; Steyer & Eid, 2001), Latent State-Trait Theory (LST theory; Steyer, Schmitt, & Eid, 1999), or Item Response Theory (IRT; Hambleton & Swaminathan, 1989; Kubinger, 1998; Rost, 2004; Steyer & Eid, 2001), for instance. In CTT, latent variables are defined as linear functions of the true scores of manifest variables. In IRT, the manifest variables are categorical or ordinal items (such as rating scales) and the latent variables are functions of the individual probabilities to answer in a given category.

Manifest covariates are important if they determine the probabilities of being exposed to intervention j. For example, *achievement scores* of students may determine the probability whether or not they take extra lessons. Sometimes, however, it is not the manifest score, but a felt deficit determining the individual intervention probability. In this case *latent* covariates measured by manifest subjective ratings are presumably more appropriate. In other applications, we may need both, manifest and latent covariates, before we can assume conditional independence of units treatments. As an example consider a *weight reduction* program. In this case, both the manifest *weight* but also the latent *psychological strain* measured by manifest ratings may determine the probability to take the weight reduction program.

Individual intervention probability functions. We may also consider the *individual intervention probability functions* $\varphi \equiv P(X = 1 \mid U)$ as covariates. They are functions of the observational-unit variable U and it can be shown that the regression $E(Y \mid X, \varphi)$ is *always unbiased*. In *true experiments*, these individual intervention probabilities may be chosen in order to obey ethical or other requirements: Patients with a high need may be assigned to intervention 1 with a high probability, whereas patients with a low need may be assigned to $X = 1$ only with a low probability. If the assignment is still random according to the individual intervention probabilities $P(X = 1 \mid U = u)$ *fixed by the experimenter*, we will still be able to identify the average causal effect of the original outcome variable Y using φ as a covariate [see Eq. (22)].

Propensities. In *quasi-experiments*, the individual, unit-specific intervention probabilities $P(X = 1 \mid U = u)$ will be unknown. This is what distinguishes a quasi-experiment from the true experiment in between-group designs. Nevertheless, if we can assume that units and treatments are independent given the covariate Z [see Eq. (25)], then the regression $E(Y \mid X, \pi)$ will also be unbiased, where $\pi \equiv P(X = 1 \mid Z)$ is called the "propensity" (Rosenbaum & Rubin, 1983). Note that Z might be a multidimensional covariate $Z = (Z1,..., ZQ)$. Then, for a large number Q of covariates in the vector Z, the regressions $E_X = {}_j(Y \mid \pi)$ within the two intervention conditions will be much simpler than the regressions $E_X = {}_j(Y \mid Z)$ of the outcomes on the original covariate Z.

Summary and conclusions

This contribution is certainly far too short to give a complete picture of what is important for experimental and quasi-experimental intervention studies. Nevertheless, some important issues have been treated, while others had to be neglected.

What we showed. Using a version of the Simpson paradox, we demonstrated that comparing means can lead to completely contradictory and misleading causal inferences unless strong assumptions can be made. In order to be able to explicate these assumptions, we then introduced the concepts of individual and average causal effects and studied how they are related to the means estimated in between-group studies. We also showed how to identify the average causal effects by empirically estimable parameters and explicated the assumptions under which identification of the average causal effects are possible. These assumptions can be secured in *true experiments*, i.e. in studies with (a) at least two intervention conditions and in which

(b) the experimenter knows the individual probabilities that a unit is exposed to intervention j. The assumptions under which identification of causal effects is possible may also hold in quasi-experimental studies if appropriate the covariates are included. These covariates might be manifest or latent. If known, the covariates can also be the probability functions $P(X = j \mid U)$ for exposure to intervention j. Otherwise, one may also use the propensities $P(X = j \mid Z)$ as covariates provided that we can assume conditional independence of units and treatments given Z.

What we did not show. We restricted our presentation of the theory to the concepts of individual and average causal effects. We treated the conditional causal effects *given the value of a covariate*. However, there are also other conditional effects which are of interest in some applications, such as the 'average effect of the treated' or the 'average effect of the untreated' (see Steyer et al., 2007, for more details). All these kinds of average causal effects provide specific information that might be of interest in evaluation studies depending on the predominant research questions. Furthermore, there are also other adjustment procedures that are not based on Equation (23), such as 'weighting the outcome variable' by a function of $P(X = j \mid U)$. We also did not treat any statistical techniques of estimating parameters and testing hypotheses nor computer programs that can be used for these purposes such as Effect Lite (Steyer & Partchev, 2007).

What should be emphasized. Most important to emphasize is that *causal inferences can most safely be drawn in true experiments*. There, unbiasedness of the regression $E(Y \mid X)$ of the outcome variable Y on the intervention variable X and/or unbiasedness of the regression $E(Y \mid X, Z)$ of Y on X and a covariate Z are secured by design, i.e., by some form of random assignment of units to interventions. Note, however, that such a design can be jeopardized by systematic attrition of the observational units (see, e.g., Shadish et al., 2002). In this case, the intervention study is not a randomized experiment any more, be-

cause (conditional) independence of units and treatments [see Eqs. (21) and (25)] is violated.

It should be noted that, in a quasi-experiment, there are two kinds of threats to the validity of causal inference. *First*, the covariate Z may not be chosen such that one of the sufficient conditions for unbiasedness (e.g., conditional independence of units and intervention conditions) holds. *Second*, even if the covariate is chosen correctly, the regression $E(Y \mid X, Z)$ may not be modeled correctly. If, e.g., one considers only a linear function of X and Z, but the true regression is a nonlinear function of X and Z , the causal inference may not be valid. Unfortunately, inference based on a covariate is not *necessarily* more valid than inference ignoring covariates. Steyer et al. (2007) provide an example in which $E(Y \mid X)$ is unbiased whereas $E(Y \mid X , Z)$ is not.

Unbiasedness of the regressions $E(Y \mid X)$ or $E(Y \mid X, Z)$ is not empirically testable. What is testable, however, are some of the sufficient conditions for unbiasedness. Conducting such 'causality tests' can contribute to protecting us from invalid causal inferences. Equation (25), e.g., implies

$$P(X = j \mid Z) = P(X = j \mid Z ,W) \qquad (26)$$

for each function $W = f (U)$ of the observational-unit variable U such as W = *gender* or W = *educational status*. If we find, for instance, that the conditional probability $P(X = j \mid Z = z, W = male)$ differs from the corresponding conditional probability $P(X = j \mid Z = z, W = female)$, we can conclude that Equation (25) does not hold. Although this finding would not *imply* that the regression $E(Y \mid X , Z)$ is biased, it would *challenge* the assumption that $E(Y \mid X , Z)$ is unbiased. This unbiasedness, however, is necessary for a valid causal inference (see Section 3.2). In this case we would consider also the confounder W to be one of the covariates and then look for other potential confounders.

Obviously, observing the covariates determining the probabilities that a unit is exposed to intervention j is of crucial importance in quasi-experiments. Without the correct covariates, the study is useless or even misleading when it comes to causal inferences. Choosing the appropriate covariates is not just statistical routine; it needs substantive knowledge about the process of assigning subjects to intervention conditions. Whenever this knowledge is insecure and whenever feasible, the true experiment should be preferred over the quasi-experiment. Otherwise, the intervention effects will remain unknown or are even estimated erroneously.

References

Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation*. Chicago: Rand McNally.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.

Holland, P. W. (1986). Statistics and causal influence. *Journal of the American Statistical Association, 81*, 945-960.

Kubinger, K. D. (Ed.). (1989). *Moderne Testtheorie*. Weinheim – Basel: Beltz.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

Neyman, J. (1923 / 1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science, 5(4)*, 465–480.

Rosenbaum, P. R. (2002). *Observational studies*. New York: Springer.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70(1)*, 41–55.

Rost, J. (2004). *Testtheorie, Testkonstruktion* (2nd ed.). Bern: Huber.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66(5)*, 688–701.

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghlin-Mifflin.

Simpson, E. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B, 13*, 238-241.

Steyer, R. (2001). Classical test theory. In C. Ragin & T. Cook (Eds.), *International Encyclopedia of the Social and Behavioural Sciences. Logic of inquiry and research design* (pp. 481–520). Pergamon, Oxford.

Steyer, R., & Eid, M. (2001). *Messen und Testen*. Berlin: Springer.

Steyer, R., & Partchev, I. (2006). *Effect Lite for LISREL*.

Steyer, R., Partchev, I., Kröhne, U., & Nagengast, B. (2007). *Causal effects in between group experiments and quasi-experiments: Theory*.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389–408.

West, S., Biesanz, J., & Pitts, S. (2000). Causal inference and generalization in •eld settings: Experimental and quasi-experimental designs. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (p. 40-84). New York: Cambridge University Press.

Winship, C., & Morgan, S. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25*, 659–707.

Yule, G. U. (1903). Notes on the theory of association of attributes of statistics. *Biometrika, 2*, 121–134.

Knowledge Management Systems: A future perspective

Marc Rittberger, Deputy Director of the German Institute for International Educational Research (DIPF)/Frankfurt

Introduction

In the late 1960s the first visionaries began to realise that information and knowledge would increasingly affect the diverse spheres of public, private and economic life. [Machlup 1962; Haiashi 1969; Bell 1973]. The transition to a knowledge society was dramatic in its speed, and mainly driven by technical developments. In the early 1990s, discussions regarding the information and knowledge society began at the political level. In the United States, this discussion was initiated in 1993 by Al Gore, Vice President at the time, and by the National Information Infrastructure Act. In Europe, the discussion was triggered in 1994 by the so-called Bangemann Report issued by the then EU Commissioner for Economy, especially Information and Media, Martin Bangemann[8]. Apart from this development, which was grounded in stronger computer networking as well as higher speed capacities for dat a transfer, the construction of an information society for the area of science had already begun in Germany in the 1970s [BMFT 1975]. This programme, too, was a direct reaction to the activities performed in the United States as a consequence from the *Sputnik Shock* in the 1950s. In this context, the Weinberg Report[9], which was published by commission of the US President in 1963, is regarded as a focal document. Even though the First Information and Documentation Programme [BMFT 1975] intended the development of professional information centres for the areas of human science and education, the development of a central information structure in Germany was at first limited to natural and technical sciences. This discussion was only re-launched for the field of education in the late 1980s, and finally the Information Center for Education (ICE) was founded at the German Institute for International Educational Research (DIPF).

The Information Center for Education at the DIPF develops modern information services and coordinates them with many partners, as well as offering these services via the Internet. At the ICE, a responsible handling of information and knowledge lies at the heart of the development of information services.

Knowledge and Information

While a detailed reference to the discussion regarding the terms ”knowledge“ and “information“, which is held in various disciplines, is not intended here, we would like to keep to the formula that is accepted in German Information Science [Krause 1998] i.e.: “Information means Knowledge in Action “[Kuhlen 1995], 34 pp., which delivers a concise and comprehensible definition of the relation between the concepts – information is knowledge that is relevant to action. Following this idea, information is grounded in a transformation process and results from knowledge following a process of “refinement”

8 In May 1994, the European Council published a report it had requested from a group of “prominent persons” on Europe and the global Information Society – the so-called Bangemann Report.

9 Report by the President's Science Advisory Committee. Panel Chaired by Alvin Weinberg: Science, Government and Information - The Responsibilities of the Technical Community and the Government in the Transfer of Information. "The Weinberg Report", 1963

[Kuhlen 1995], p. 34, which results in the creation of "added informational values". In this respect, [Kuhlen 1999], p.138 distinguishes knowledge and information by explaining that "knowledge means immaterial cognitive structures as verified statements (to certain degrees) regarding objects and circumstances that pertain to a fictional or real world."10

Knowledge needs to be accessible and transferable in order to be communicated and presented in a manner that it can be received. The representation of knowledge in machine-based information systems is realised by data that are stored in the information system. The concept suggested by common language practice that knowledge itself is stored in information systems such as books or databases is wrong.

> "Of course, books as such do not contain any knowledge, and neither do machine-based information systems store knowledge. Rather, they represent knowledge by means of defined syntactic codes, that is data, that are interpreted according to a set of agreements/rules, which themselves can be put into pragmatic contexts by agreements/rules. Knowledge can only be re-activated from data if a person makes the effort to understand the data. Accordingly, a book cannot contain any knowledge as it does not bear any mechanisms for reading and understanding its own data."[11]

The determination of knowledge defines its truth, following Euler [Kanngießer 1984], knowledge can be determined by experience, logics or communication. The determinations themselves can be very different, and we can accordingly outline an epistemological knowledge scale for the truth value of information terminology. The distinction between true and false does not pertain to information, which is instead characterised by reliability, use, relevance for action, topical value, completeness and costs [Kuhlen 1995], p.40.

Following these reflections, we would like to first comment on the term, and current trends of knowledge management and discuss the terms of knowledge communication and quality. Furthermore, the Information Center for Education of the DIPF will be introduced, which regards itself as a national competence centre for educational information in the Federal Republic of Germany. Its most important services will be briefly introduced and the Information Center activities will be described with regard to the transfer of knowledge from educational research to the general public as well as to policy-makers.

Knowledge Management

A meaningful and productive use of knowledge constitutes one of the objectives of knowledge management. Many contexts of knowledge management focus on the shop-floor situation of an enterprise. In most cases these centre on a holistic perspective that includes all kinds of human-technically oriented interventions and measures as far as they are adequate for optimising the production, reproduction, distribution or use or logistics of knowledge within an organisation [Schüppel 1997]. [Bullinger et al. 1998] and [Probst et al. 1999] describe knowledge management from the perspective of enterprise. An understanding of knowledge management that is less focused on business enterprises can be found in [Reinmann-Rothmeier & Mandl 1997], p.22 (quoted from [Lehner 2006], p.33),

10 "Mit Wissen sind immaterielle kognitive Strukturen als (in unterschiedlichem Ausmaß) gesicherte Aussagen über Objekte und Sachverhalte der (fiktiven oder realen) Welt gemeint."

11 "Natürlich enthalten Bücher für sich kein Wissen, und auch maschinelle Informationssysteme für sich enthalten kein Wissen, sondern sie repräsentieren nur Wissen über syntaktisch definierte Zeichen, also Daten, die wiederum durch Konventionen/Regeln semantisch interpretiert und ebenfalls durch Konventionen/Regeln in pragmatische Zusammenhänge gestellt werden können. Wissen kann erst wieder aus Daten aktiviert werden, wenn jemand es unternimmt, die Daten zu verstehen. Ein Buch kann deshalb nichts wissen, weil es keine Mechanismen enthält, seine eigenen Daten zu lesen und zu verstehen." [Kuhlen 1995], p.39

where knowledge management is described as a challenge to society, an organisational task, which renders individual and social competence necessary. They name the following important tasks and abilities:

- Disseminating information,
- Selecting and assessing information,
- Contextualising information and rendering it meaningful
- Constructing knowledge from information, and developing new knowledge
- Associating knowledge and building knowledge networks
- Saving, structuring and updating knowledge
- Passing on knowledge, mediating and distributing knowledge,
- Exchanging knowledge and complementing one another's knowledge,
- Using and implementing knowledge,
- Assessing knowledge-based action and developing new knowledge accordingly.

Considering this or similar bullet lists that provide closer descriptions of the concepts underlying knowledge management, the comparison with concepts of information transfer and information development is noticeable. The tasks and abilities listed by [Reinmann-Rothmeier & Mandl 1997] are regarded as necessary for the development and provision of successful information services, too. Both perspectives focus on communication of information and knowledge as a core element. We will take a closer look at two aspects of knowledge management in the following paragraphs, that is knowledge communication and the quality of knowledge applications.

Knowledge Communication

Communication as a focal aspect of knowledge can either be synchronical, or asynchronous. To communicate knowledge describes an effort to exchange knowledge, „the (in most cases) intentional interactive construction and mediation of insights and capacities at both a verbal and a non-verbal level" [Eppler & Reinhardt 2004], p.2. The communication of knowledge requires a sender and a recipient interacting with each other, who can define a shared context in order to render the communication and the transferred knowledge meaningful. [Eppler & Reinhardt 2004], p.3 distinguish between the communication of knowledge and the communication of information, which trigger the strong contextualisation, the explanatory contents, the question of truthfulness of knowledge, the necessity of legitimising and the motivation as well as the understanding.

There are different ways of exchanging knowledge. From the perspective of research, two types of exchange by means of technical processes are particularly interesting. The visualisation of knowledge and communicating knowledge in groups. An overview in terms of a periodic table[12] regarding important methods of visualising knowledge communication is categorised by methods that describe data, information, concepts, strategies and metaphors of visualisation as well as mixtures of the methods mentioned here [Lengler & Eppler 2007]. The categorisation of different methods of visualisation for knowledge communication enables us to find methods of visualisation for similar purposes quickly.

Another area of knowledge communication concerns applications in the context of Social Software, e.g. Wikis and Blogs which support the exchange of texts, but also services such as "YouTube[13]" or flickr[14] , where the exchange of images or videos is supported.

12 http://www.visual-literacy.org/periodic_table/periodic_table.html
13 http://www.youtube.com/
14 http://www.flickr.com/

Here, the World Wide Web serves as a platform where contents are no longer provided by large publishing houses or television corporations, but where individual Internet users are the providers. Web 2.0 is characterised by content-driven applications, a strong networking structure, a participation of the users, simple marketing models, continuous software development and the combination of software [O'Reilly 2005]. These framework conditions allow for the development of communities that pursue the same interests and where people jointly work at projects in a collaborative, distributed manner. The most well-known example of such a collaborative project is the online encyclopedia Wikipedia[15], where anybody who feels competent can contribute articles or improve those articles others have written.

Quality and Trust in Knowledge

How good is the knowledge, can I rely on the information, what criteria speak in favour of the quality of knowledge? Contrary to information quality, [Rittberger 2004a] the quality of knowledge can only be described in the context of terms such as true, complete, certain, unambiguous and trustworthy. Since an assessment of the quality of knowledge always includes a transfer, or communication of knowledge, the value of knowledge always depends on the competence and capacity of the bearer of the knowledge, the faultlessness of the transfer and the capacities of the recipient. The criterion of trustworthiness/reliability of knowledge seems to be particularly relevant to us. [Giddens 1997], 48 pp. describes trust in a list of ten components such as the trust that is derived from complete information, or the trust that is related to contingency and reliability, but also the trust resulting from a belief in the reliability of persons or systems.

[Rittberger 2004b] describes that trust in information systems and in the knowledge they transfer is closely related to reliability, credibility, correctness, competence and transparency. Hence, information providers must strive to fulfil these criteria in order to communicate high value knowledge by their information systems.

The most important information services offered by the Information Center for Education are described in the following paragraph. Two short examples will illustrate the areas where, within the context of information services offered by the ICE, we have started some projects to enhance knowledge communication and undertaken measures to improve the quality of the information services we provide.

Information Center for Education

The information services offered by the Information Center for Education rest on two pillars, that is the German Education Portal (Fachportal Pädagogik) and the German Education Server (Deutscher Bildungsserver).

Starting from a loose cooperation of thirty pedagogical documentation institutions in Germany, which were integrated into the model project Educational Information System (EIS) (Fachinformationssystem (FIS) Bildung [Botte 1997]), the EIS was developed as a reference system of literature in the field of national and international pedagogical science and educational research, which documents and disseminates pertinent references by the German Education Index (FIS Bildung Literaturdatenbank). A total of approximately 640.000 documents are registered with an annual increase rate of 25,000 documents. The focus of the reference system concerns literature published in German. Since 1999, the German Education Index has been accessible on the Internet.

15 http://de.wikipedia.org/wiki/Hauptseite

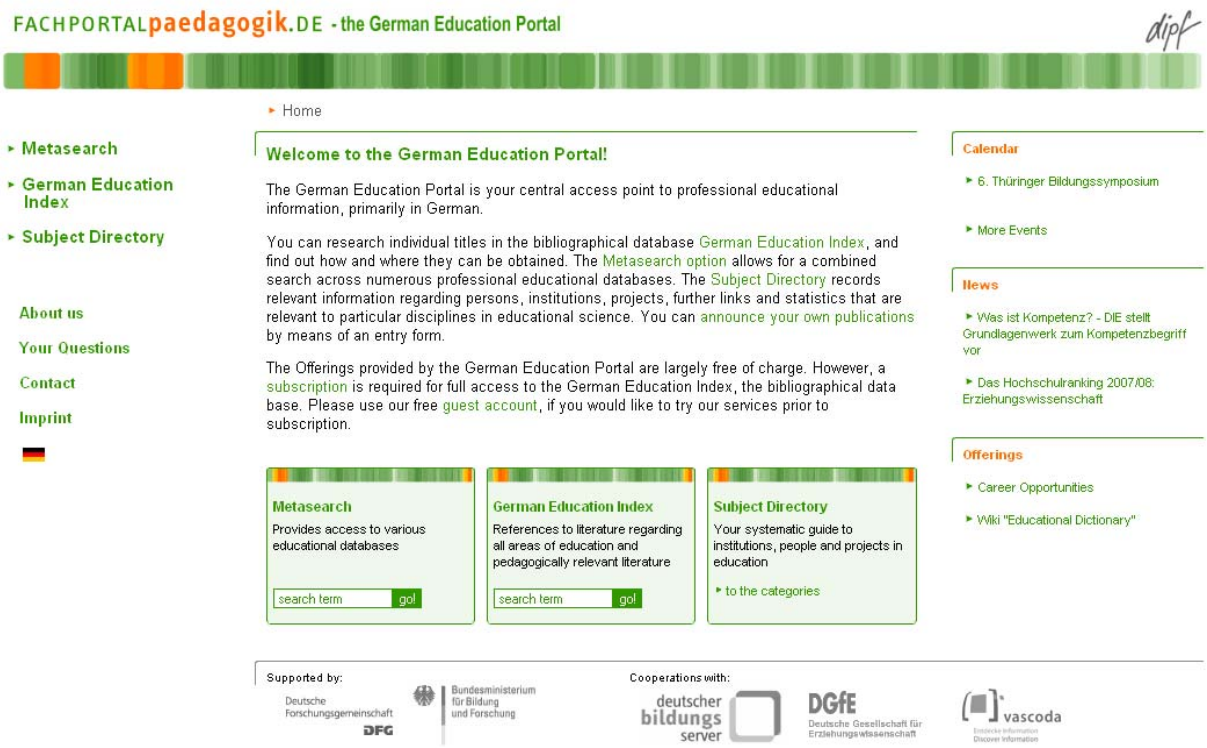


Figure1*: The German Education Portal.*

The information institutions in Germany have responded to the growing interdisciplinary or trans-disciplinary work in science and research, and to the network-oriented work and research processes. Two directions can be discerned which on the one hand, concern a national networking of relevant information institutions for each discipline, and on the other hand, provisions for a central point from where resources can be accessed at a national level. To this purpose, the German Education Portal was established as a central access point for the field of education/pedagogics[16] [Kunz 2006] which was launched on the Internet in the summer of 2005.

Apart from the German Education Index, the Portal provides resources on institutions, persons and projects from different areas of education. It is further possible to conduct thematic searches in other databases, such as the British Education Index, in databases of events, in educationally relevant articles published in 39 national and international newspapers, or in newsletters. Hence, the German Education Portal offers researchers from the area of educational science a central access point to educational scientific information, including full text provisions, which is intended to support them in their research and lecturing activities.

16 http://www.fachportal-paedagogik.de/start_e.html

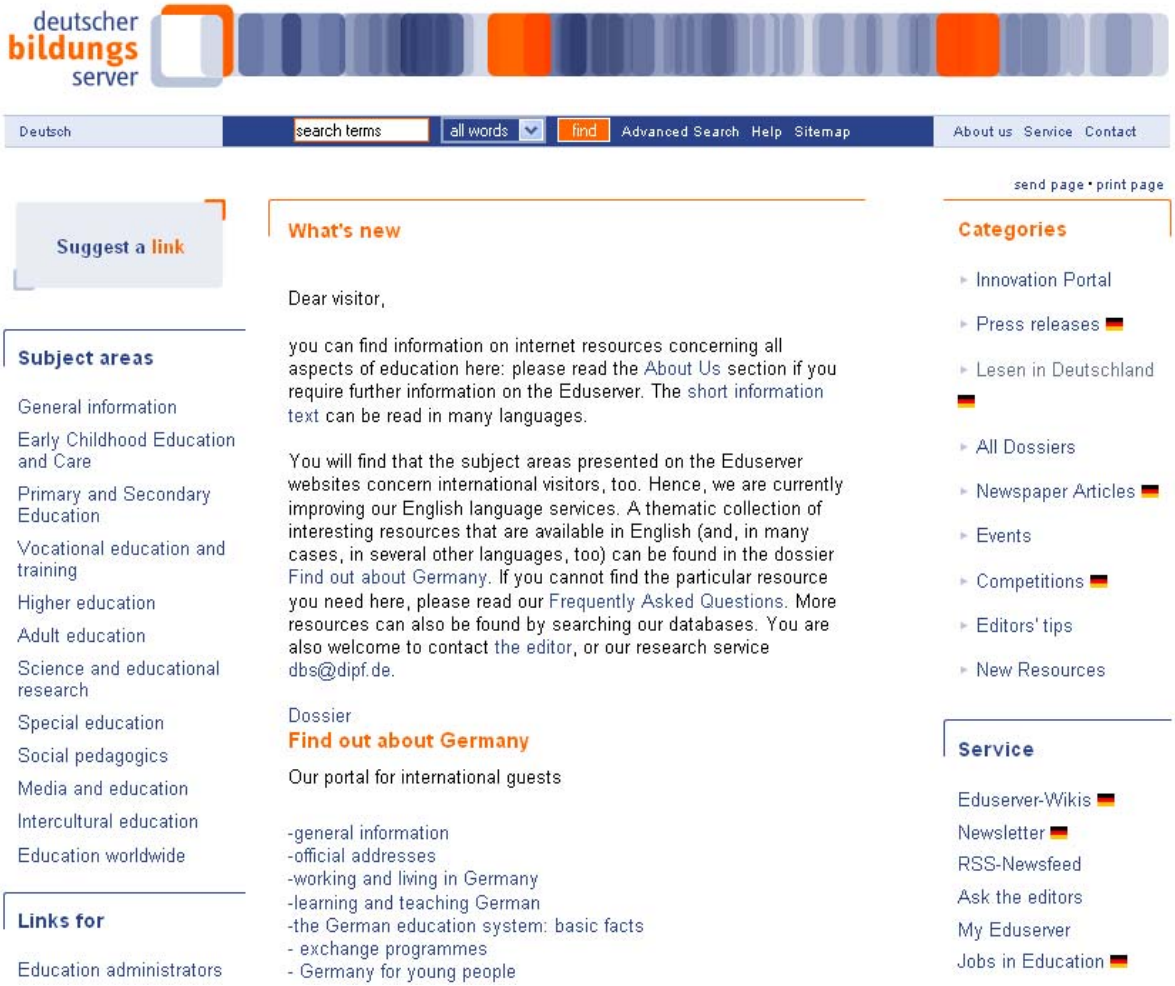


Figure 2*: The German Education Server (Eduserver)*

In addition to the scientific clientele, the educational sector addresses a number of groups who are active either as users, or as providers, of educational services. Further groups include administrators, policy-makers, parents, media, and the general public with an interest in education. The German Education Server[17] (DBS) [DIPF 2005], which was conceptualised as early as the mid 1990s at the Humboldt University in Berlin, has become the most important, most popular and at the same time the most complex information service provided by the Information Center for Education. Its services meet the demands of modern information systems in the World Wide Web in the best possible way, and it corresponds to the targets set out by the level of politics and its programmes in Germany [BMBF 2002; BMWA & BMBF 2003].

The German Education Server addresses the needs of its users [Botte 2005] by

- Offering diverse collections of data and information for all areas and levels of education,
- Offering access to mainly full text information, which is available in electronic form,
- Paying particular attention to official documents published by the federal government and the Länder states, and describing their contents,

17 http://www.eduserver.de/

- Structuring access to information by user groups and subject areas.

Besides a number of other portals and databases, and the ongoing development of existing services both from its own resources and in project-funded research and development work, the Information Center for Education has enhanced its services and focus by means of two projects funded by the Federal Ministry of Education and Research (BMBF), which are particularly relevant to German educational research in an international context:

- Office for International Cooperation in Educational Research[18], which supports German empirical educational science in the attempt to internationalise. Therefore, the Office counsels and supports international research projects, initiates international financing strategies, contributes to the dissemination of scientific research findings, and finally mediates and supports the use of national research findings in an international setting.
- Technology Based Assessment[19], a project for the development of a service and research structure to support the technical and psychometric enhancement of technology based assessment methods.

Summary

The Information Center for Education at the German Institute for International Educational Research offers high quality, knowledge intensive services. To maintain the high standard, evaluations, user studies and usability assessments are carried out on a regular basis (for example, in co-operation with the higher education institutions of Darmstadt, Stuttgart or Potsdam). Internal developments, for instance towards optimising the process of database construction, are decomposed into the part processes of selecting resources, indexing, abstracting and conversion of data and subsequently analysed. These regular assessments and evaluations, which are performed by external partners, enable the Information Center for Education to continue developing its services following an impartial and critical survey, and to improve them with a close eye to its clients. Since June 2007, the German Education Server Wikis[20] are offered within the context of web 2.0. In addition to a Wiki on world literature and a Wiki-based information exchange, the Education Wiki, which is a Wikipedia fork, addresses all users who are interested in education, in order to build a comprehensive educational WIKI encyclopedia together with its users, and enhance its quality by Social Tagging, that is shared indexing. The Information Center for Education in Germany looks at itself as a partner for information, communication and knowledge transfer in the triangle of research, practice, and politics & administration which describes the partners of an evidence based policy. We offer services for the aforementioned partners in education to overcome communication disabilities with tools for communicating knowledge on a high information quality standard.[21]

18 http://interkoop.dipf.de/

19 http://www.tba.dipf.de/

20 http://www.bildungsserver.de/zeigen.html?seite=5277

21 Many thanks to Gwendolyn Schulte for translating this contribution.

References

Bell, D. (1973). The coming of post-industrial society. New York: Basic Books

BMBF (2002). Strategisches Positionspapier. Information vernetzen - Wissen aktivieren. http://www.bmbf.de/pub/information_vernetzen-wissen_aktivieren.pdf

BMFT (1975). Programm der Bundesregierung zur Förderung der Information und Dokumentation (IuD-Programm) 1974-1977.

BMWA & BMBF (2003). Informationsgesellschaft Deutschland 2006. Aktionsprogramm der Bundesregierung. http://www.bmbf.de/pub/aktionsprogramm_informationsgesellschaft_2006.pdf

Botte, A. (1997). Fachinformationssystem Bildung. Bericht über den Modellversuch der Bund-Länder-Kommission "Erprobung der Konzeption für ein Fachinformationssystem Bildung." Frankfurt am Main: DIPF

Botte, A. (2005). Der deutsche Bildungsserver als Antwort auf aktuelle Anforderungen an die Fachinformation Bildung. dipf informiert. Fünf Jahre - Geschäftsstelle Deutscher Bildungsserver Nr.1, 5-8

Bullinger, H.-J., Warschat, J., Prieto, J., und Wörner, K. (1998). Wissensmanagement - Anspruch und Wirklichkeit: Ergebnisse einer Unternehmensstudie in Deutschland In: Information Management, Jg. 1998, H. 1, S. 7-23. IM Information Management Nr.1, 7-23

DIPF (2005). 5 Jahre - Geschäftstelle Deutscher Bildungsserver. DIPF informiert, Sonderheft, http://www.dipf.de/publikationen/dipf_informiert_sonderheft_dbs.pdf

Eppler, M. J. & Reinhardt, R. (2004). Zur Einführung: Das Konzept der Wissenskommunikation. In: Wissenskommunikation in Organisationen. Methoden, Instrumente, Theorien. Reinhardt, R. & Eppler, M. J. (eds.); Berlin: Springer, 1-12

Haiashi, Y. (1969). Johoka Shakai. Tokyo

Kanngießer, S. (1984). Simulationskonzepte des Wissens und der Grammatik. Niemeyer: Tübingen, Vol. C.-R. Rollinger (Hrsg.): Probleme des (Text-)Verstehens. Ansätze der Künstlichen Intelligenz. Symbolik[]Sprache und Information Band 10Symbolik[], 22-44

Krause, J. (1998). Alles schon da? - Der handlungsorientierte Informationsbegriff als Grundlage informationswissenschaftlicher Theoriebildung und Gestaltungsarbeit. Ethik und Sozialwissenschaften. Streitforum für Erwägungskultur, 9 Nr.2, 223-225

Kuhlen, R. (1995). Informationsmarkt. Chancen und Risiken der Kommerzialisierung von Wissen. Universitätsverlag Konstanz: Konstanz, Schriften zur Informationswissenschaft

Kuhlen, R. (1999). Die Konsequenzen von Informationsassistenten. Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden? Frankfurt/Main: Suhrkamp, suhrkamp taschenbuch wissenschaft

Kunz, N. (2006). Das Fachportal Pädagogik des Deutschen Instituts für Internationale Pädagogische Forschung. Bibliotheksdienst, 40 Nr.12, 1393-1405

Lehner, F. (2006). Wissensmanagement. Grundlagen, Methoden und technische Unterstützung. München: Hanser

Lengler, R. & Eppler, M. (2007). Towards A Periodic Table of Visualization Methods for Management. Datum: 2007,

http://www.visual-literacy.org/periodic_table/periodic_table.pdf

Machlup, F. (1962). The production and distribution of knowledge in the United States. Princeton, NJ: Princeton University Press

O'Reilly, T. (2005). What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. Datum: 2005, http://www.oreillynet.com/pub/a/oreilly/ tim/news/2005/09/30/what-is-web-20.html

Probst, G. J. B., Raub, S. und Romhardt, K. (1999). Wissen managen. Wie Unternehmen ihre wertvollste Ressource optimal nutzen. Frankfurt am Main: Frankfurter Allgemeine Zeitung für Deutschland

Reinmann-Rothmeier, G. & Mandl, H. (1997). Wissensmanagement. Phänomene - Analyse - Forschung - Bildung. München: Ludwig-Maximilians-Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie, Forschungsbericht Nr. 83

Rittberger, M. (2004a). Informationsqualität. In: Grundlagen der praktischen Information und Dokumentation. Handbuch zur Einführung in die Informationswissenschaft und -praxis. Kuhlen, R., Seeger, T. und Strauch, D. (eds.); München: Saur, Kap. B 17, 315-321

Rittberger, M. (2004b). Vertrauen und Qualität in Informationsdienste. Wo finde ich Vertrauen im Information Quality Framework? In: Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen. Hammwöhner, R., Rittberger, M. und Semar, W. (eds.); Konstanz: UVK Verlagsgesellschaft, 153-165

Schüppel, J. (1997). Wissensmanagement. Organisatorisches Lernen im Spannungsfeld von Wissens- und Lernbarrieren. Wiesbaden: Deutscher Universitäts-Verlag, Gabler-Edition Wissenschaft

Use of Evidence: Development of an Evaluation Culture in Education

Ritva Jakku-Sihvonen, Director, Ph. D., Quality Assurance and Monitoring, Helsinki, Finland

Development of education and teaching should be based on the findings of national and international evaluations and research. Close co-operation between researchers, administrators, policy-makers, teachers and other practitioners is key in successful development of the education system and teaching arrangements. This kind of knowledge-based policy-making needs genuine knowledge in order to facilitate reforms that are really useful for students. (Jakku-Sihvonen & Niem 2007) Furthermore, knowledge-based development needs teachers and civil servants with competence to analyse and reflect on the validity of knowledge as well as to ask creative and critical questions.

The function of evaluation has increased in terms of public administration. According to a report by the Finnish Parliamentary Committee on Education in 1998 (3/1998 – HE 86/1997):

> "Evaluations are to support continuous development of education to facilitate improved learning, as well as realisation of objectives set for schooling and learning by various sections of legislation. The Parliamentary Committee wishes to underline that evaluation also has an important social and political function in enhancing the realisation of equality among people within the Finnish education system."

Evaluations are typically used to know the outcomes of certain activity or to understand how relevant, cost-effective, efficient or effective certain activity or program is. Furthermore, evaluative research can produce new knowledge for developing education.

General remarks

Evaluation generally means making an interpretative analysis of the phenomena or activities being studied, and defining the benefit or value produced by these activities. Evaluation typically aims to uncover findings that are as valid as possible. Evaluation must be based on relevant information sources and transparent criteria. The methods used for analysis of data have to be recorded and justified. In addition, evaluations made to support or change education policies have to be done by evaluators with a high level of expertise and high professional ethics. Evaluative statements and conclusions must be based on evidence. These criteria lay the foundation for evaluating the state of affairs: their strengths and weaknesses, benefits and disadvantages.

Evaluative conclusions are important. Evaluative conclusions are meant to crystallise the main findings of the evaluation. There are also two main schools of thought in terms of recommendations based on evaluative conclusions made by evaluators. Some evaluators are willing to add their recommendations to the evaluation reports and others want to leave recommendations to administrators and policy-makers. Programme evaluators are more interested in making recommendations while researchers are more interested in leaving recommendations to the policy-makers.

The main similarities and differences between evaluation and research (see, for instance, Popham 1988, 13) are as follows:

- Knowledge based on evaluation is
  - produced by scientific methods
  - valid and often context-bound
  - aimed at defining the value of the action or phenomena
- Knowledge based on research is
  - produced by scientific methods
  - valid, reliable and general
  - aimed at generating new knowledge

Evaluation can serve at different levels: individuals, classrooms or other kind of groups, institutional, local, regional, national, European and global level. Depending on the purpose of evaluation, methods can vary: there are large-scale assessments, based on samples or a complete set of material, thematic evaluations, programme evaluations, evaluations of state, evaluative surveys, different kinds of self-evaluations, internal and external quality assurance systems, auditing etc.

It is always important for evaluators to work independently: evaluators and researchers need to be autonomous and free to perform their work without any pressure. The ethic of evaluators has been reflected on significantly. There are some general principles that evaluators have to follow. For instance, in Finland, the following principles have been accepted:

- The aim of evaluation is to yield information to serve development of education.
- Information sources and analysis methods used in evaluation are to be recorded and justified.
- Evaluative statements and conclusions are based on relevant quantitative and/or qualitative data and analysis.
- The evaluator's familiarity with the object being evaluated, as well as his/her ethics and understanding of the nature of the human and social effects produced by education are good enough.
- As society is democratic, society is pluralistic, the evaluations demand a lot of the evaluators in terms of their understanding of the social, national and international contexts that they are working in.
- All partners (teachers, students etc.) have to be informed about the purpose of the evaluation.
- The evaluator has to work independently. He/she is responsible for the validity of the evaluation.
- (Framework for Evaluating Education in Finland 1999)

What kinds of evaluations are useful?

To make evaluations useful, the purpose of the evaluation has to be clear enough for all partners and stakeholders. The evaluator's familiarity with the object being evaluated, as well as his/her ethics and understanding of the nature of the human and social effects produced by education are important. The aim of evaluation is to yield information to serve development of education.

There are two main ways of using educational evaluations and research in policy-making:

1. To support decisions made previously.
2. To act as a foundation for new political decisions.

In both cases, policy-making and policy have to be democratic and transparent and have to be useful in terms of equality of education. To support knowledge-based policy-making between European countries and work, the following principles should be taken into account:

- All partners can use their voices and are aware of their own responsibilities.
- Evaluators must have independent and autonomous positions.
- Evaluators are responsible for the validity of the findings and for making evaluative conclusions.
- Decision-making based on evaluation has to be transparent.
- Politicians and other decision-makers are responsible for decisions made based on evaluations.
- There is a need for European and international co-operation.

In recent years, evaluation of education has been progressing in all sectors of public administration. In the field of education, many kinds of evaluation have been implemented. By means of evaluation, we wish to serve the needs of political and administrative guidance by producing information which is useful and applicable in terms of development of education. When evaluating educational outcomes, the essential task is to produce information that is up-to-date and relevant in terms of educational development at all levels. Local administration and schools produce evaluation data for their own development activities. School-based evaluations – both pure self-assessments and involving external consultants – started to become more common during the 1980's. Analyses are drafted on the schools' or institutes' strengths and weaknesses and serve as a basis for various development and consultancy projects.

In the 1990s, accountability thinking and different kinds of programme evaluations became very popular. At the same time, the concepts used for public administrative evaluation also became well-known among evaluators of education. In Finland, for example, a general framework for evaluating educational outcomes was accepted and concepts like effectiveness, efficiency and economy were defined for evaluating the education system. Based on the system for evaluating outcomes (cross outcome effect) used in state administration in general, concepts were defined for educational administration as follows:

Education has been successful when objectives set nationally and internationally for each organisational form of the education system, for each school or institution and for the learning activities of each individual, have been achieved.

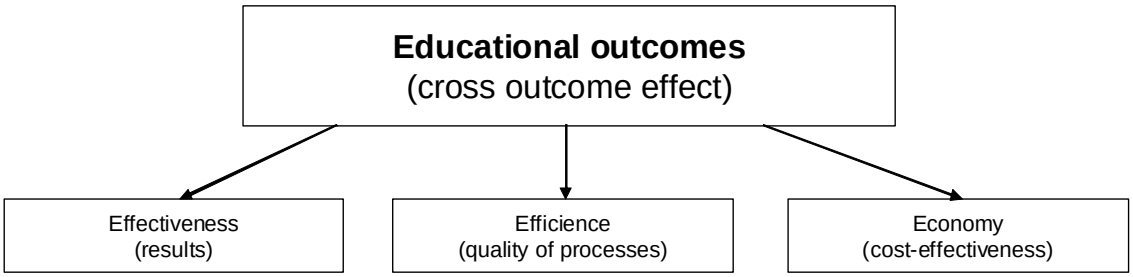


Figure. *Conceptual model of evaluating educational outcomes.*

Education is efficient when the functionality, flexibility and timing of the education system, educational administration and teaching arrangements are as appropriate as possible and when the quality of teaching is high.

Education is effective when it succeeds in preparing people, both qualitatively and quantitatively, promoting the mental growth of individuals and development of society, culture and working life.

Education is economical when resources within education have been allocated in an optimal way with respect for educational objectives and when the amount of resources is in proportion to the number of educational services provided and appropriate in terms of the structure and organisation of service provision.

Evaluations can be voluntary or based on statutes or agreements. In Finland, for example, new educational legislation came into force at the beginning of 1999 (the Act on basic education, chapter 5, section 21) and defines educational evaluation as follows:

> "The purpose of educational evaluation is to ensure that the purpose of this law is manifested in practical terms and to support development of education and to improve conditions for learning. The key thing for schools is to evaluate instruction provided as well as its effectiveness and also to participate in external evaluations of educational activities".

At the moment, almost all important European and national policy-programmes are evaluated. In educational branch also large-scale international assessments (like IAE, TIMMS, PISA, IALS etc.) have been implemented. The results of these comparative analyses can be used to develop indicators for national, European and international policy-making and political purposes. They can also be used for development of curricula.

Many kinds of quality assurance systems have been developed specifically for vocational and higher education. Creating a common European Higher Education Area to improve the competitiveness and attractiveness of European higher education has increased the need for promotion of European co-operation in quality assurance with a view to developing comparable criteria and methodologies. (See Standards and Guidelines for Quality Assurance in the European Higher Education Area 2007) Auditing quality assurance systems at universities and polytechnics has increased. Quality-oriented thinking has been stimulating self-evaluation activities and has prepared the ground for analyses of customer satisfaction. This quality-oriented thinking has involved stimulating self-evaluation activities in schools and has paved the way for analyses of customer satisfaction and for the growth of an evaluation culture in the field of education in general. Results of these evaluations are mainly used to develop administrative processes and learning conditions.

Two positive examples

Developing Research-based Teacher Education

An interesting example concerning the use of evaluations is Finnish teacher education, which has been evaluated many times over the past 15 years. Reform of teacher education in Finland began in the 1970's as the result of close co-operation between administrators and professors. The motive for reforming teaching was to support the new comprehensive school, which was aimed at supporting the policy of equality in education. A political decision was made to transfer teacher education to universities. The following principles to develop masters' level teacher education were accepted:

- All teacher education for comprehensive and upper-secondary schools should be academic and carried out at universities.
- Teacher education should be unified for different teacher categories.

- Initial education for future teachers must give a common and broad qualification to all teachers and this common background can then be flexibly complemented by in-service education.
- Pedagogical studies should be developed in such a way that teachers are prepared to be educators in the broad sense of this concept and can attend to their pupils' socio-emotional growth. Teachers should have a pedagogical, optimistic attitude to their work that is based on the latest research. *Theoretical and practical studies as well as subject academic matter and pedagogical studies should be integrated more successfully.*
- Teacher education should consist of societal and educational policy studies. (Committee report, 1975).

Developing research-based teacher education is a continuous project. In order to develop teacher education, many national and international evaluations have been carried out. During the last 15 years, 7 evaluations have been implemented, 3 of them have been provided internationally by international evaluation groups. (Niemi & Jakku-Sihvonen 2006, 32-33.)

Evaluations have indicated that the research component plays an essential role in teacher education. Professors Westbury, Hansen, Kansanen and Björkvist (2005) have described the function of the science of Education in class teacher education as follows:

> "The aim is a balanced programme in which research-based pedagogical thinking is the central organising concept across the three major strands in pre-service studies: subject didactics, educational theory and teaching practice. These areas are seen as being in reciprocal interaction, with the organising principle being the conviction that there should be a research base for practice...The function of the science of education is essential in teacher education. By emphasising the importance of good knowledge in research methodology used in the science of education and sufficient substance knowledge students are prepared for critical thinking and autonomous decision-making...The function of the writing of the master's thesis education prepares us to use and analyse the research results."

At the moment, all teachers in Finland have to complete a Master's degree (300ECTS). In terms of the new Bologna process, this degree is equivalent to the second cycle degree in the European higher education area. Research-based teacher education and well-organised in-service training are very important preconditions for teachers' ability to utilise the results of evaluations and research in school life. On the other hand, the findings based on evaluations of teacher education help teacher educators to develop the curricula of teacher education.

KivaKoulu –programme to prevent bullying at schools

KivaKoulu –research (2007-2009) is an ongoing empirical study on practices preventing bullying at schools. The project is implemented by the University of Turku and financed by the Ministry of Education. Within the project there are about 4000 pupils, who are 10 to 12 years old. There is a randomized experiment group and a control group. This project is aimed to find new research-based working practices.

The project is lead by professor: *Christina Salmivalli*. For more information, please visit http://www.kivakoulu.utu.fi/pages/yhteydenotot.html

Need for cooperation

If the goal is to develop a knowledge-based professional culture in the educational arena, we need a research–based working culture among professionals as well as a new kind of

administrative culture. To support the preconditions for knowledge-based working culture in education, Professor Hannele Niemi (2007) made the following suggestions to increase the application of knowledge in developing education:

- European level research programmes
    - teaching, school administration, educational policy and teacher education in focus.
- Action research programmes
    - as part of teachers' and other professionals' continuous education.
- National and European doctoral programmes
    - for teachers, headmasters and teacher educators (three cycles of the Bologna process).
- Local and national multi-professional networking
    - researchers, teachers, social-workers, working-life partners and media.
- A European project "Evaluation and evaluation culture"
    - towards more participatory and communicative culture. Active interaction between researchers, practitioners and policy-makers is needed.

To conclude

Evaluating accountability has become very popular. However, there is also a need for evaluations serving developments of education and evaluative research producing new knowledge. When developing the education system, teachers, teacher educators and educational administrators are in key positions. In addition, researchers and evaluators need resources not only to draft valid reports but also as part of the dissemination process.

Close co-operation between policy-makers and evaluators is possible. To achieve successful co-operation, we need researchers and evaluators who are not only critical and analytical but who will be willing to help make education more human and supportive and we need policy-makers and professionals who are willing and able to listen to researchers and understand the value of new findings.

References

Committee report, 1975. (in Finnish) The Ministry of Education. Helsinki. Finland.

A Framework for Evaluating Educational Outcomes in Finland. 1999. Helsinki:National Board of Education.

Jakku-Sihvonen,R.&Niemi,H. 2007. Introduction. In Jakku-Sihvonen, R. & Niemi,H. (eds.): Education as a Societal contributor. Reflections by Finnish Educationalists. Frankfurt a.M.: Peter Lang, 9-20.

Niemi, H.& Jakku-Sihvonen, R. 2006. Research-based Teacher Education. In Jakku-Sihvonen, R. & Niemi,H. (eds.) Research-based Teacher Education in Finland –Reflections by Finnish Teacher Educators. Helsinki: Finnish Educational Research Association, 31–50.

Niemi, H. 2007. Unpublished presentation.

Popham, W.J.1988.2nd ed. Educational evaluation. New Jersey:Prentice Hall.

Standards and Guidelines for Quality Assurance in the European Higher Education Area 2007.Helsinki:European Association for Quality Assurance in Higher Education. (http://www.enqa.eu/pubs.lasso)

Westbury, I., Hansen, S.- E., Kansanen, P. & Björkvist, O. 2005. Teacher Education for Research-based Practice in Expanded Roles: Finland's experience. Scandinavian Journal of Educational Research 49(5), 475–485.

Evaluating education and labour market interventions: The search for the counterfactual

John Elliott, Department for Education and Skills, United Kingdom

Introduction

This paper presents an overview of two methods for policy evaluation and illustrates their use with examples from the UK. This paper accompanies a presentation given by John Elliott[22] at the 'Knowledge for Action' conference in Frankfurt, 28-30 March 2007.

Policy evaluation – the search for the counterfactual

One of the key challenges with policy evaluation is to determine what would otherwise have happened

When undertaking policy evaluation we want to measure how a treatment 'D' affects some outcome of interest 'Y', with the added complication that 'Y' is also affected by many other influences 'X'.

Ideally we would want to know whether the outcome having received the treatment $Y_{1i} = f(X_i, D=1)$ is different from the outcome had the treatment not taken place, $Y_{0i} = f(X_i, D=0)$. The problem is that we cannot observe both outcomes for the same individual.

Standard econometric techniques will only return a good estimate of the impact of a treatment if strong restrictions are met

Faced with such a problem a standard approach is to estimate an Ordinary Least Squares (OLS) equation[23].

Estimation using OLS: $Y_i = \alpha' X_i + \beta D_i + u_i$

However, this would only generate a good estimate of 'β' – the impact of the treatment on the outcome – if, for example:

- There is common support: that is the characteristics – the X's – of the treatment group overlap with that of the control group;
- The X's are not affected by the treatment;
- The X's are evenly distributed across the two groups; and
- There are no unobservable characteristics which influence the outcome that are not common across the two groups – i.e. 'β' is unaffected by omitted variable bias.

As can be seen these are strong restrictions so in practice OLS estimation will often not estimate 'β' with sufficient accuracy.

22 John Elliott is the Chief Economist at the Department for Education and Skills.

23 Similar arguments would apply to other standard econometric models – for example if the outcome was a di chotomous variable and a probit or logit equation was estimated.

Random Assignment and Propensity Score Matching are two approaches that can address these problems

A number of different techniques can be used to overcome the inadequacies of OLS. This paper focuses on two methodologies, Propensity Score Matching (PSM) and randomised experiments.

PSM, like OLS, is a non-experimental matching technique. To be used there must be both a treatment group and an appropriate non-treatment group. A control group is created by selecting individuals from the non-treatment group that match members of the treatment group. Rather than matching on all characteristics, individuals are matched on their propensity score – the probability of being in the treatment group given their characteristics.

In a *randomised experiment* people are randomly assigned to treatment and control groups. If these groups are large enough, the distribution of all pre-treatment characteristics should be identical in both groups so any difference in the outcome can be attributed to the treatment.

The following sections will illustrate both techniques using their application in a UK based policy evaluation[24].

Evaluation of educational maintenance allowance

Educational Maintenance Allowance was introduced to raise participation, retention and achievement in post-compulsory education

During the late 1990's, whilst the attainment of UK school leavers continued to rise the proportion participating in non-compulsory education remained relatively flat. In addition, international comparisons showed that participation in education at age 17 was below the OECD average. The Educational Maintenance Allowance (EMA) was devised to encourage more young people to remain in post compulsory education.

EMA provides an income-tested allowance paid directly to 16-18 year olds who remain in full-time education. Rates in 05/06:

- Income up to £20,270 per annum - £30 per week
- Income of £20,271 to £24,850 per annum - £20 per week
- Income of £24,851 to £30,000 per annum - £10 per week

EMA was introduced into 15 pilot areas in the 1999/2000 academic year and rolled out to a further 36 areas a year later. National roll-out for 16 year olds took place in 2004/05 and in subsequent years for 17 and 18 year olds.

PSM was used in the evaluation of EMA

For the evaluation of EMA 10 pilot areas (plus Leeds and 4 areas of London) were matched with 11 control areas. Two cohorts of 16 year olds were followed for four years – the first cohort were 16 in 1999/2000 whilst the second were 16 in 2000/01. Four variants of the

24 A more in depth discussion of EMA and its evaluation can be found in: http://www.dfes.gov.uk/research/data/uploadfiles/RR678.pdf. A more detailed discussion of the initial ERA evaluation is contained in: http://www.dwp.gov.uk/asd/asd5/rports2007-2008/rrep412.pdf

policy were tested – weekly payments to young person or their family, retention and achievement bonuses.

Although chosen to be similar, pilot and control areas could still be quite different in demographic comparisons that may be relevant for participation decisions. Therefore, PSM involved matching at an individual level rather than doing comparisons at area levels.

Many characteristics were used for matching including; household income, the young person's earlier educational outcomes, ward-level measures of post-16 participation (ward is small geographical area used for local elections) and measures of school quality. The key assumption is that there are no other factors relevant to participation that differ significantly between pilot and control areas.

The evaluation found that EMA had a significant impact on participation

The evaluation found that EMA had a significant impact on participation rates. For example, amongst eligible Year 13 pupils[25], EMA increased participation by 7.3 percentage points. As one would expect given EMA is income targeted, the greatest impact was on those from lower Socio-Economic Groups.

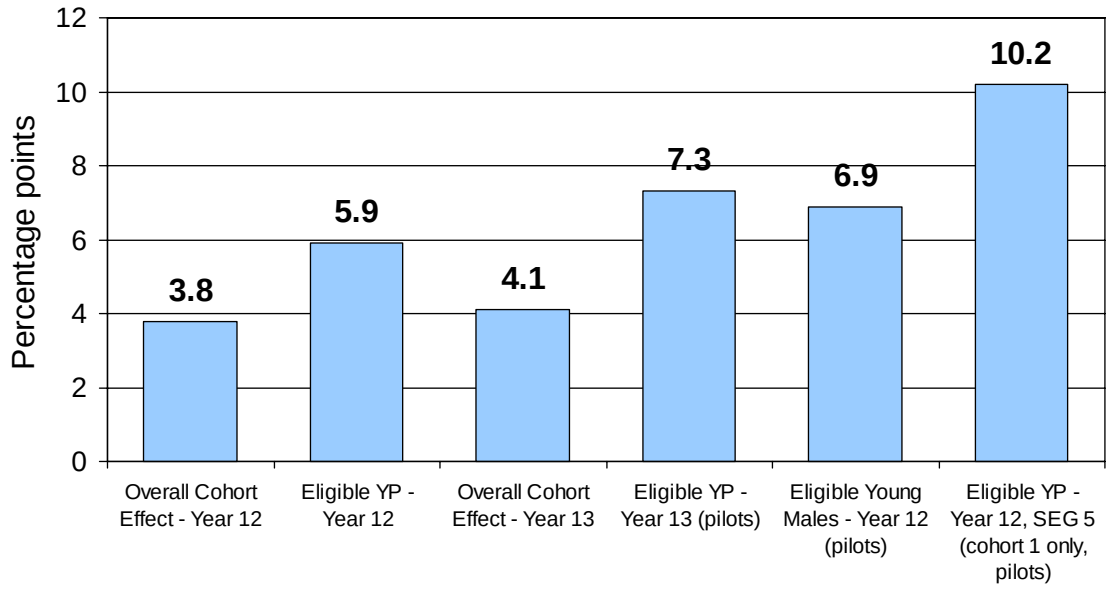


Chart: *The percentage point change in participation resulting from the introduction of EMA in pilot areas*

Evaluation of employment retention and advancement

The Employment Retention and Advancement policy was seen as a 'next step' in British welfare-to-work policies

Employment Retention and Advancement (ERA) was launched in 2003, envisioned as a 'next step' in British welfare-to-work policies. It follows on from successful active labour market policies, such as the New Deal for Young People, which have helped reduce unemployment.

25 Academic age 17.

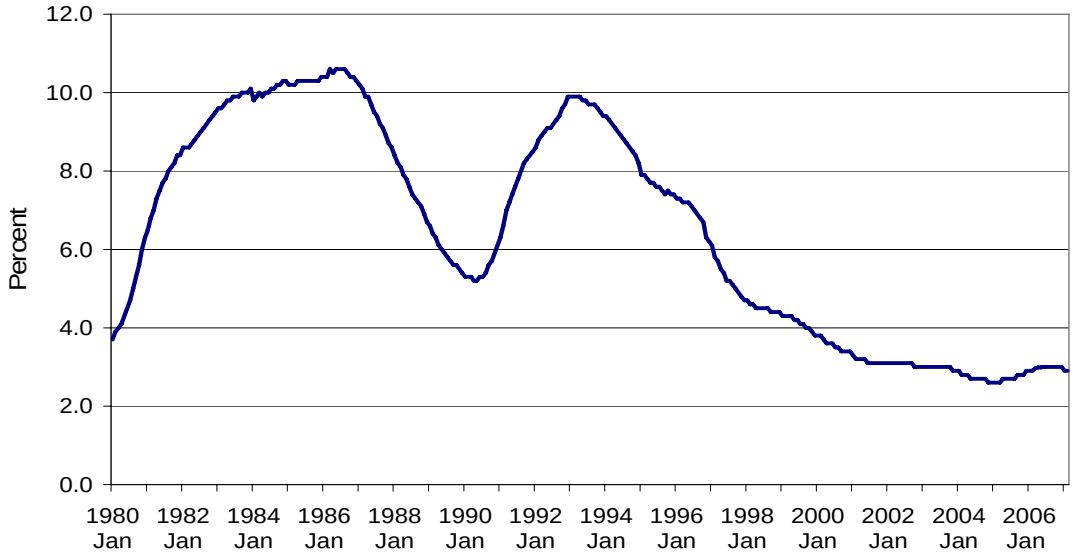


Chart*: UK unemployment – claimant count rate*

The aim of ERA is to help individuals to sustain employment and progress in work. To do this ERA offers a combination of job coaching and financial incentives once participants find work. For example, those in the programme group will receive individual support to assist them into suitable work and then to remain and advance in work. They may also receive cash incentives for remaining in full time employment as well as payments to support training and access to emergency funds to overcome short-term barriers to retaining work.

ERA was piloted in six UK regions – East Midlands, London, North East England, North West England, Scotland and Wales. Within these pilot areas ERA was targeted at three client groups:

- Unemployed lone parents;
- Lone parents in work but on low wages – who are in receipt of the Working Families Tax Credit; and
- The long term unemployed aged at least 25

ERA was evaluated using a random assignment experiment, with volunteers randomly assigned to treatment and control groups

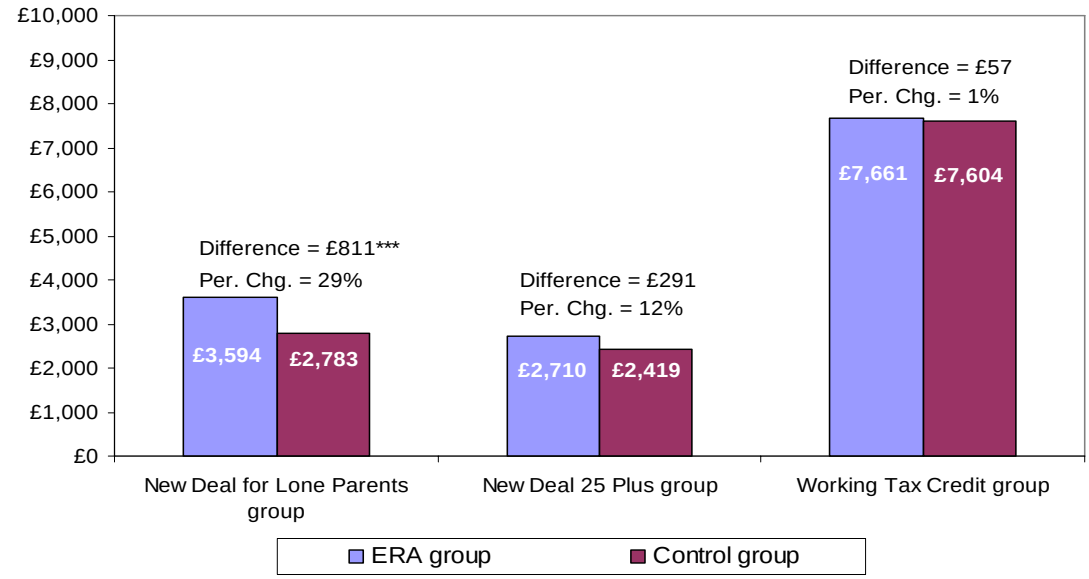
Members of the three qualifying groups were invited to volunteer for ERA. Whilst refusals were low, there was a significant number of people within the relevant client groups who were not selected to participate in the pilots. The implications of this for the evaluation of ERA are discussed in a recent report[26].

A total of 16,000 volunteers were randomly assigned to control and treatment groups. This large sample means that the characteristics are evenly distributed across the two groups. As such any differences in the average outcomes between control and treatment groups can be attributed to ERA.

26 Goodman and Sianesi (2007) 'Non-participation in the Employment Retention and Advancement Study: A quantitative descriptive analysis'. http://www.dwp.gov.uk/asd/asd5/WP39.pdf

The evaluation suggests ERA had a positive impact on both earnings and employment

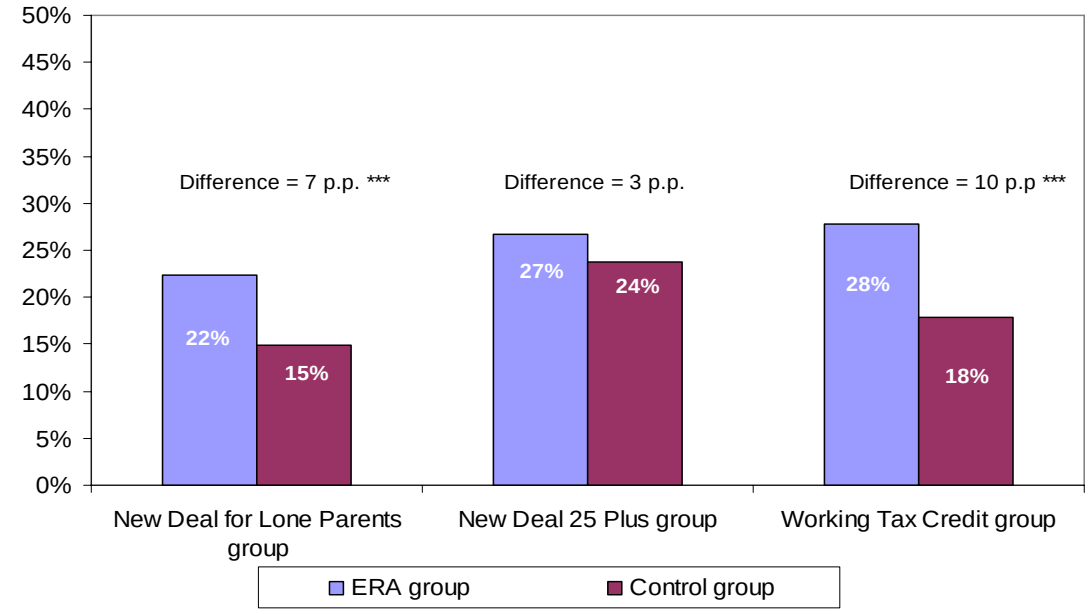
The results of the initial ERA evaluation suggest that the earnings of the treatment groups was higher for each of the qualifying groups. However, the difference is only significant for unemployed lone parents who earned on average 29% more than the control group.



Statistical significance: *** = 1 per cent

Chart: *ERA's first year impacts on average earnings by client group*

Much of the increase in earnings amongst the unemployed lone parents is attributed to ERA making them more likely to be in work and when in work to be working full time. The following chart shows that that likelihood of being in full time employment was higher for each customer group – for two this difference was significant.



Statistical significance: *** = 1 per cent

Chart: *ERA's impacts on the likelihood of working full time at the end of year 1, by customer group*

Discussion

So is random assignment preferable to matching in a technical sense?

Random assignment is often see as the 'gold standard'. If the population is large enough, and assignment truly random, then there will be common support; an even distribution of the Xs; and no selection on unobservables.

Given this the estimated impact of the treatment on the treated group can be assumed identical to the potential impact of the treatment on the control group – but not necessarily the remaining population[27].

There can also be practical advantages to random assignment, for example it can be relatively cheap – as much less data needs to be collected than is the case for matching methods which require information on the Xs.

However, if the population is not large, or assignment not truly random then matching methods may be preferable as:

- They allow direct tests for the degree of common support and the distribution of the Xs, even if they may be susceptible to the impact of unobservables.
- They can take account of different distributions of observables between the treatment and control groups. So the estimated impact of the treatment on the non-treated needn't be the same as the effect of the treatment on the treated.

27 For example, the impact of ERA on those who didn't volunteer for the pilots may be different to those who did in such a way that the impact of the treatment is altered. Similarly, the characteristics of pilot areas may differ from other geographical areas.

Operational and political factors can also shape the way policies are to be evaluated

Beyond the statistical estimation of the impact the problems with random assignment are:

- The association with a 'lottery' – especially with what may be an unrepeated event for individuals.
- Operationalising the random assignment – for example delivering two or more policies through the same institutions and ensuring boundaries between treatment and control groups are clear[28].

There are of course also problems with PSM

- The cost – collecting data on all X's many of which will be beyond existing sources such as administrative data
- The need to work across boundaries when the pilot and control areas are geographic.

28 For a more detailed discussion see http://www.dwp.gov.uk/asd/asd5/WP39.pdf and http://www.dwp.gov.uk/asd/asd5/rports2005-2006/rrep265.pdf

Annex

Empfangsrede

Jutta Ebeling, Bürgermeisterin der Stadt Frankfurt, im Kaisersaal des Römer
Empfang der Stadt Frankfurt am 28.03.2007

Sehr geehrte Damen und Herren,

Als Bürgermeisterin der Stadt Frankfurt ist es eine große Ehre, dass eine wichtige Bildungs-Konferenz im Rahmen der deutschen EU-Ratspräsidentschaft hier in Frankfurt stattfindet. Wir befinden uns hier im Kaisersaal des Frankfurter Römer, der "gudd Stubb", wie die Frankfurter/innen sagen. In diesem Saal versammelten sich seit dem Mittelalter die Kurfürsten des Reiches, wenn sie nach Frankfurt kamen, um einen neuen Kaiser zu wählen. Der eigentliche Wahlakt und die Huldigung fand dann im benachbarten Dom statt. 1838, ein Jahrzehnt vor der Paulskirchenversammlung träumte man in Frankfurt von einer geeinten Nation und setzte ihr in diesem Saal ein Denkmal. Die 52 Kaiserbildnisse wurden von den damals renommiertesten Zeitgenössischen Künstlern erstellt. Alle 52 Kaiser, angefangen von Karl dem Großen (Nr. 1) bis zu Kaiser Franz dem II (Nr. 52) sind hier zu sehen.

Meine Damen und Herren,

mit dem Ende des Heiligen Römischen Reiches Deutscher Nation endete nicht die Geschichte des Kaisersaals. 1846 etwa fand hier die erste Versammlung der deutschen Germanisten statt, bei der Jacob Grimm, der Begründer der deutschen Sprachwissenschaft, auf die geschichtliche Bedeutung dieses Raumes hinwies.

Bis zum heutigen Tage finden hier immer wieder wichtige Tagungen und Zusammenkünfte statt, und viele von Ihnen können sich noch an die Willkommens-Feier der deutschen Fußball-Nationalmannschaft nach der Fußball Europameisterschaft erinnern, als unsere Fußball-Kaiser auf dem Balkon nebenan die Huldigungen des Volkes auf dem Römer entgegennahmen.

Meine Damen und Herren,

Ich wollte die Evidenz und der Konsens in Bildungsfragen glichen dem des König Fußball.

Lassen Sie mich als Bildungsdezernentin dieser Stadt ein paar wenige Bemerkungen zum Thema Ihrer Tagung machen:

50 Jahre sind seit der Unterzeichnung der Römischen Verträge vergangen. In den vergangenen fünf Jahrzehnten wurde Beispielloses erreicht: Wirtschaftliches Wachstum, Stabilität, Wohlstand für viele ein Leben in Frieden und Freiheit sind nur einige wenige Stichpunkte.

Der europäische Einigungsprozess ist ein Glücksfall.

Allerdings befindet sich die EU zur Zeit in einer Phase des Nachdenkens. Viele BürgerInnen haben Angst, dass die Union sie durch Bürokratisierung und Bürgerferne einschnürt und überfordert. Deshalb bedarf das historische Projekt der EU eines neuen demonstrativen Schubs, denn wir alle wissen; Zur Europäischen Union gibt es keine Alternative. Entscheidend ist, wie wir in der EU mit Bildung und Forschung umgehen. Sie sind die Quelle des künftigen Wohlstands, denn sie sind Voraussetzung für Wachstum und Beschäftigung. Der Schlüssel für die Zukunftschancen jedes einzelnen aber auch der Gesellschaft liegt in Bildung, Ausbildung und Weiterbildung. Daher ist es mehr denn je an der Zeit, unsere Kräfte, unsere Kreativität und unsere Ressourcen zu vereinen und durch gemeinsame Anstrengungen in Bildung und Forschung zu einer positiven Entwicklung für Europa beizutragen.

Bei PISA hat Deutschland nicht weltmeisterlich abgeschlossen, denn nicht zuletzt Pisa hat gezeigt, dass der Schulerfolg eines Schülers oder einer Schülerin in keinem Land in Europa so stark von der sozialen Herkunft abhängt wie bei uns in Deutschland (insb. Bayern). Das (Pisa) Beispiel anderer Länder zeigt, dass mehr Chancengleichheit möglich ist und dass es an uns liegt, wenn wir sie nicht erreichen.

In deutschen Schulen wird das Bildungskapital, das junge Menschen von zu Hause mitbringen, viel zu häufig nur verwaltet. In anderen Ländern haben junge Menschen stärker die Möglichkeit, ihre Talente zu entfalten und verborgene oder zu Hause nicht geförderte Begabungen zu entdecken. Das muss auch in Deutschland möglich werden; die Schulen müssen zu einem Teil das ausgleichen, was junge Menschen von zu Hause nicht mitbringen.

Das ist die wichtigste Aufgabe, die unsere Schulen in den kommenden Jahren erfüllen müssen. Die Lehrkräfte müssen sich um jeden einzelnen kümmern, weil Bildung eine wesentliche Voraussetzung dafür ist, dass das eigene Leben gelingt: Sie ist eine Voraussetzung für persönliches Glück und sie ist deshalb auch eine wichtige Voraussetzung dafür, dass sich der einzelne als Teil einer Gemeinschaft und für die Gesellschaft verantwortlich fühlt. Deutschland braucht die bessere Förderung jedes einzelnen, weil fehlende Chancengleichheit den gesellschaftlichen Zusammenhalt und auf Dauer den sozialen Frieden gefährdet.

Wir brauchen das auch, weil unsere Wirtschaft in den kommenden Jahren und Jahrzehnten mehr denn je auf qualifizierte Frauen und Männer angewiesen sind, damit deutsche Unternehmen erfolgreich sein können. Soziale Verpflichtung und wirtschaftliche Notwendigkeit weisen beide in die gleiche Richtung. Wir dürfen niemanden abschreiben, niemanden fallen lassen. Jeder junge Mensch muss individuell gefordert und gefördert werden.

Beim Bau eines Hauses beginnt man aus gutem Grund mit dem Fundament und nicht mit dem Dach. Das pädagogische Haus beginnt zunächst im Geburtshaus und in den Kindergärten. Am Tag der Einschulung in das staatlich regulierte Bildungsleben sind Kinder für die Schule vorbereitet oder auch nicht.

In Frankfurt am Main haben wir uns schon lange vor PISA intensiv mit frühkindlicher Bildung auseinandergesetzt und sind schon ein gutes Stück vorangekommen. Durch Fachtagungen mit Erzieherinnen, Sprachförderung in jeder Kita und der festen Verankerung eines "Bildungsnetzwerks KITA" ist frühkindliche Bildung in Frankfurt schon lange ein Schwerpunkt der Arbeit. Erhöhung der Personalbemessung, um Zeit für individuelle Bildungsprozesse zu haben.

Das liegt in der Verantwortung der Kommune. Hier handelt Frankfurt gestützt von dem Wissen um die Bedeutung frühkindlicher Bildung für den Bildungsprozess insgesamt. Ich habe etwas ratlos vor dem Titel ihrer Konferenz gesessen und gegrübelt, was er wohl bedeuten mag. "Wissen für Handeln." Wissengestütztes Handeln ist immer besser als blinder Aktionismus. Meinen Sie das? "Forschungsstrategien für eine evidenzbasierte Bildungspolitik" war eine noch größere Herausforderung, an der ich gescheitert bin.

Evidenz ist Augenscheinlichkeit, Einsicht, unmittelbare Gewissheit des anschaulich eingesehenen, so gegoogelt.

Augenscheinlich müssen die Bildungsvoraussetzungen in der BRD besser werden. Evident ist, dass Länder mit längerem gemeinsamen Lernen bessere Erfolge erzielen. Richtig ist, dass Ganztagsschulen Schulerfolge unterstützen.

Und doch leitet das, was mir evident erscheint, nicht durchgängig die Bildungspolitische Strategie in der Republik. Nicht zuletzt deshalb weil Bildungspolitik oft ein Spielball parteipolitischer Auseinandersetzung und institutioneller Zersplitterung ist.

In den Bundesländern werden mit Bildungspolitik Wahlen verloren. Die Kommunale Ebene von Bildungsverantwortung wird ganz selten mitgedacht, anders als im Vorzeigeland Finnland z.B.

Die Föderalismusreform, die die Entscheidungskompetenz ausschließlich den Ländern zugeschrieben hat, verstärkt aus meiner Sicht die kleinstaatliche Zersplitterung und damit den ideologischen Einfluss, wo gemeinsame nationale Anstrengung nötig wäre, verhindert damit einen breiten gesellschaftlichen Konsens, den es braucht, um Deutschland in der Weltklasse bezüglich Bildung spielen zu lassen. Um in der Sprache des Fußballs zu bleiben: Wir stellen keine erfolgreiche Nationalelf. Das unterscheidet Bildungspolitik vom Fußball – leider.

Wenn Sie in dieser Tagung mit dem im eigentlichen Wortsinn fragwürdigen Titel Forschungsstrategien entwickeln, die so evident sind, dass keine Politiker/innen an ihnen vorbeikommt, dann sind wir alle einen großen Schritt weitergekommen.

Meine Damen und Herren, ich wünsche Ihnen in diesem Sinne eine erfolgreiche Fortsetzung der Konferenz und hoffe, dass einige von Ihnen die Zeit für die Schönheiten Frankfurts außerhalb des Konferenzsaals finden werden. Den Verantwortlichen des Bundesministeriums für Bildung und Forschung und der Europäischen Kommission danke ich für die Ausrichtung dieser Konferenz in Frankfurt. Als Mitglied des Stiftungsrates des Deutschen Instituts für Internationale Pädagogische Forschung ist es mir eine besondere Freude, dass die Organisation der Konferenz diesem Institut übertragen wurde. Ich danke den Verantwortlichen für die Durchführung und wünsche allen Teilnehmern einen angenehmen Abend und eine anregende Konferenz.

Empfangsrede

Kornelia Haugg, Leiterin der Abteilung „Berufliche Bildung; Lebenslanges Lernen“ des BMBF.
Empfang der Stadt Frankfurt am Main im Kaisersaal des Römer am 28.03.2007 19.30

Sehr geehrte Damen und Herren,

die turnusmäßige Organisation der Ratspräsidentschaft bringt es mit sich, dass eine Präsidentschaft für ein Land ein „seltenes“ Ereignis ist – und einige werden sagen „Gott sei Dank“. Die Wichtigkeit und internationale Sichtbarkeit dieses Ereignisses bedeuten aber auch, dass die Platzierung eines Themas auf der Agenda der Ratspräsidentschaft in der Regel eine starke politische Botschaft darstellt – nach innen (national), wie nach außen (europäisch, international). Dies gilt auch für diese Konferenz und das damit verbundene Thema.

Deutschland, genauso wie viele andere Staaten, hat in den letzten 10-15 Jahren eine starke Wandlung durchgemacht, was die Rezeption und Verarbeitung von Ergebnissen empirischer Bildungsforschung anbelangt. Wer sich an die teilweise doch recht emotionalen Diskussionen um TIMSS erinnert (gerade in Deutschland), wird wissen, dass diese Diskussionen - wenn überhaupt - eine sehr geringe Wirkung auf Realpolitik hatten. Und derjenige wird mir wiederum zustimmen, dass sich durch PISA und die nachfolgenden nationalen wie internationalen Anstrengungen in dieser Beziehung viel geändert hat.

Trotzdem wäre es falsch, sich mit dem Erreichten zufrieden zu geben. Im Gegenteil: Verantwortungsvolle Forschungspolitik versucht positive Ergebnisse nachhaltig zu festigen, sowie bestehende Schwächen zu beseitigen. Das Bundesministerium für Bildung und Forschung stellt sich dieser Verantwortung. Diese Konferenz ist dabei nur ein Baustein in einer weiterreichenden Anstrengung auf dem Weg zu wissenschaftlich begründbarem Entscheiden und Handeln in der Bildungspolitik.

Unser „keynote speaker“, Prof. Levin, der wie kaum ein anderer beide „Seiten“ kennt - die Bildungsforschung und die Politik – hat uns in seinem Vortrag die Komplexität der Problematik aufgezeigt. Er hat aber auch an wichtigen Punkten klar Stellung bezogen und aufgezeigt, wie Lösungen aussehen können, was bereits funktioniert bzw. was eben nicht funktioniert. Der Vortrag und die aufgezeigten Lösungen stimmen mich optimistisch, dass der vom BMBF eingeschlagene Weg zielführend und richtig ist. Ich möchte mich auch heute Abend noch einmal ausdrücklich bei Herrn Levin bedanken, dass er sich die Zeit für diese Konferenz genommen hat. Sein beruflicher Werdegang führt ihn gerade von der Politik zurück in die Wissenschaft, wo er schwerpunktmäßig mit dem Thema unserer Konferenz befasst sein wird. Ich wünsche ihm auch in dieser „neuen alten“ Rolle weiterhin viel Erfolg.

Die Erfahrungsberichte aus Dänemark und aus den Niederlanden bestätigen meine bisherige positive Wahrnehmung. Sie zeigen aber auch: Wir müssen dazu lernen, und ein zentraler Baustein dieses Lernprozesses wird internationale Kooperation sein. Ich danke Herrn Bugge Bertramsen und Herrn Stegeman für ihre Ausführungen. Sie zeigen auch auf, wie ein solcher Lernprozess aussehen könnte.

Im Kontext internationaler Kooperation – auch und gerade in diesem Bereich – wird die Zusammenarbeit mit der Europäischen Kommission in Zukunft noch wichtiger werden. Heute danke ich Frau Quintin und ihren Mitarbeitern für die Unterstützung bei der Vorbereitung und Durchführung dieser Konferenz und für die erkennbare Bereitschaft der Kommission, den eingeschlagenen Weg fortzusetzen.

Meine sehr geehrten Damen und Herren: ich bin überzeugt, Sie teilen meine Einschätzung, dass wir heute einen überaus interessanten Konferenzauftakt erlebt haben. Ich wünsche Ihnen eine gute Fortsetzung der Konferenz, anregende und konstruktive Diskussionen, Knüpfen neuer Kontakte und eine Vertiefung der schon vielfach vorhandenen Kooperationen. Für heute aber einen entspannenden Abend hier im Römer.

Grußwort

Ingrid Müller-Roosen, Bundesministerium für Bildung und Forschung
Empfangsrede im Städel in Frankfurt am Main anlässlich des Diners am 29.03.2007

Sehr geehrte Damen und Herren,

Im Namen des Bundesministeriums für Bildung und Forschung begrüße ich Sie alle hier im wunderbar dekorierten Frankfurter Städel Museum und freue mich, damit zum geselligen Teil dieses Konferenztages überleiten zu dürfen.

Und dieser Konferenztag war aus meiner Sicht ein guter Tag. Nach dem ersten, durch Vorträge geprägten Teil haben wir heute einen stärker interaktiven und kommunikativen Arbeitstag erlebt. Ich denke, es war auch ein diskussionsreicher Tag, der uns unserem Streben nach „Wissen für Handeln" wieder ein ganzes Stück näher gebracht hat.

Nicht ganz unwichtig für diese Diskussionen waren sicherlich die perfekte Vorbereitung und das angenehme Ambiente dieser Konferenz. Vielen Dank den vielen Helferinnen und Helfern und vor allem den Vertretern des DIPF dafür!

Für mich war heute wieder bemerkenswert, wie unterschiedlich – wie komplex – die einzelnen zu bedenkenden Aspekte im Kontext der evidenzbasierten Bildungspolitik sind – einmal abgesehen von den begrifflichen Schwierigkeiten. Diese Komplexität und natürlich die grundlegende Relevanz waren und sind zentrale Gründe für das Bundesministerium für Bildung und Forschung, diesem Thema eine sichtbare Rolle im Rahmen der deutschen EU-Ratspräsidentschaft einzuräumen.

Hier gilt es nun sicherlich, die bisher zusammengetragenen Erkenntnisse auch auf die einzelnen Handlungsebenen herunter zu brechen, um diese für unsere tägliche Arbeit in Zukunft noch besser nutzbar zu machen.

Wie Sie sicher wissen, ist das Bundesministerium für Bildung und Forschung bereits in der Vergangenheit maßgeblich an der Vorbereitung und Finanzierung von „large scale assessments" auf nationaler und vor allem auch auf internationaler Ebene beteiligt. Die Bereitstellung finanzieller Mittel für solche Studien ist dabei ein notwendiger Aspekt – aber ebnen nur **ein** Aspekt. In gleichem Maße müssen wir nämlich – in Kooperation mit allen Akteuren – eben **die** Rahmenbedingungen schaffen, um solche Studien in der geforderten und auch notwendigen Qualität durchzuführen.

Mein Dank geht an Prof. Tuijnman und Prof. Steyer dafür , dass sie uns heute eindrücklich vor Augen geführt haben, welcher Anstrengungen es – auch auf Seiten des Forschungsmanagements – noch bedarf, um dabei internationalen Standards zu genügen und um überhaupt die Ergebnisse, wie sie - da spreche ich in eigner Sache – die Politik so gern hätte: schnell und belastbar – produzieren zu können. Neben diesen an wissenschaftlichen Erfordernissen orientierten Überlegungen, wird es auch gelten – eine weitere Erkenntnis insbesondere des heutigen Tages – den „Übersetzungsprozess" von Wissenschaft in Politik zu verbessern.

Prof. Levin hat bereits am ersten Tag anschaulich darauf hingewiesen, dass dies nicht in einem Schritt gelingen kann, sondern dass dafür viele verschiedene Faktoren notwendig sind. Heute haben Prof. Rittberger und Herr Leney Lösungsvorschläge dargestellt und Ausblicke auf die Zukunft gegeben.

Ich würde mich freuen, wenn diese Diskussionen und Nachfolgeaktivitäten dazu auch Impulse für eine nachhaltige Forschungspolitik setzen werden um diesen Prozess künftig noch stärker zu professionalisieren.

Frau Jakku-Sihvonen und Herrn Elliott danke ich dafür, dass sie das Thema der Konferenz heute in einen breiteren Zusammenhang gestellt haben. Über den inhaltlichen Wert ihrer jeweiligen Beiträge hinaus zeigen aus meiner Sicht gerade diese beiden Vorträge auch den Wert internationaler Konferenzen und internationaler Kooperation. So liegen ja in diesen beiden Ländern – Finnland und Großbritannien – z.T. langjährige Erfahrungen für unterschiedliche Teilaspoekte des Themas unserer Konferenz vor. Die Kompetenz der Vortragenden, gepaart mit dieser spezifischen Erfahrung, ergibt einen eindeutigen Mehrwert, der in rein nationalem Kontext nur schwer zu erzeugen ist.

Vielen Dank daher in diesem Zusammenhang insbesondere all denjenigen, die auch weitere Wege nicht gescheut haben, um an dieser Konferenz teilzunehmen und sich einzubringen.

Jetzt möchte ich Ihnen – möchte ich uns allen durchaus nicht uneigennützig – nicht länger den verdienten kulinarischen Genuss vorenthalten. Und apropos kulinarischer Genuss: nicht versäumen möchte ich einen herzlichen Dank an das Land Hessen, das uns heute großzügig mit Wein versorgt.

Sehr geehrte Damen und Herren, liebe Konferenzteilnehmer: Ihnen allen danke ich im Namen des Bundesministeriums für Bildung und Forschung für Ihre aktive Teilnahme am heutigen Tag.

Ich hoffe sehr, dass sich die Kommunikation und Kooperation, die wir in den Workshops erlebt haben, in Zukunft auch über die Konferenz hinaus fortsetzen wird. Das Bundesministerium für Bildung und Forschung, sicherlich auch in enger Kooperation mit der Europäischen Kommission, wird seinen Teil dazu tun, dies auch künftig weiterhin zu ermöglichen.

Ich wünsche Ihnen allen einen angenehmen Ausklang des heutigen Tages!

Vielen Dank.