



Was bedeutet es, die Wirkung von Weiterbildung zu messen? Was kann überhaupt wie gemessen werden? Und stellen sich in der Wissenschaft, in der Praxis und bei Fördergebern unterschiedliche Fragen bezüglich der Wirkung von Bildungsprozessen? Hierüber sprach Redakteur Jan Rohwerder mit Prof. Dr. Andreas Martin, Leiter der Abteilung System und Politik am Deutschen Institut für Erwachsenenbildung (DIE) und Professor für Bildungswissenschaft an der FernUniversität in Hagen, und Dr. Sabine Schwarz, Leiterin des Bereichs Grundbildung in der Lernenden Region - Netzwerk Köln und Teil des Kompetenzzentrums GrundbildungsPFADEpfade.

Schlagworte: Erwachsenenbildung; Weiterbildung; Rezeptionsforschung; Bildungsforschung; Weiterbildungsangebot; Eingliederung; Zuwanderer; Sprachkurs; Effektivität; Weiterbildungseinrichtung; Teilnehmer; Lernerfolg; Messung

Zitiervorschlag: Martin, A. & Schwarz, S. (DIE) (Hrsg.) (2025). »Die mangelnde Fehlerkultur ist das größte Hindernis bei der Wirkungsmessung«. *weiter bilden*, 32(2), 14-18. Bielefeld: wbv Publikation. <https://www.wbv.de/artikel/WBDIE2502W005>

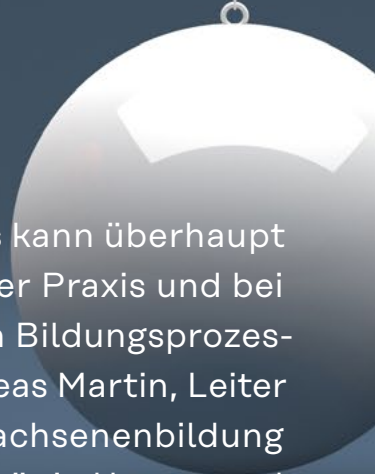
E-Journal Einzelbeitrag
von: Andreas Martin, Sabine Schwarz

»Die mangelnde Fehlerkultur ist das größte Hindernis bei der Wirkungsmessung«

aus: Wirkungen (WBDIE2502W)
Erscheinungsjahr: 2025
Seiten: 14 - 18
DOI: 10.3278/WBDIE2502W005

»Die mangelnde Fehlerkultur ist das größte Hindernis bei der Wirkungsmessung«

WEITER BILDEN spricht mit ANDREAS MARTIN und SABINE SCHWARZ



Was bedeutet es, die Wirkung von Weiterbildung zu messen? Was kann überhaupt wie gemessen werden? Und stellen sich in der Wissenschaft, in der Praxis und bei Fördergebern unterschiedliche Fragen bezüglich der Wirkung von Bildungsprozessen? Hierüber sprach Redakteur Jan Rohwerder mit Prof. Dr. Andreas Martin, Leiter der Abteilung System und Politik am Deutschen Institut für Erwachsenenbildung (DIE) und Professor für Bildungswissenschaft an der FernUniversität in Hagen, und Dr. Sabine Schwarz, Leiterin des Bereichs Grundbildung in der Lernenden Region – Netzwerk Köln und Teil des Kompetenzzentrums GrundbildungsPFADE.

WEITER BILDEN: Wenn wir über die Wirkung von Weiterbildung sprechen – was lässt sich eigentlich messen?

ANDREAS MARTIN: Grundsätzlich lässt sich vieles, ich würde sogar sagen, fast alles messen – es kommt auf die Datengrundlage, auf die verfügbare Zeit und auf die verfügbaren Ressourcen an. In der Weiterbildung hatten wir lange Zeit nur begrenzte Datengrundlagen, aber die werden inzwischen kontinuierlich breiter. Diese Daten für die Forschung nutzbar zu machen, benötigt vor allem personelle Ressourcen, und daran mangelt es oft noch. Grundsätzlich sind aber die Methoden zur Wirkungsmessung vorhanden, um theoretische Annahmen zu testen und auch, um kausale Zusammenhänge zu prüfen. Wenn es aber auch darum gehen soll, bildungspolitische Interventionen oder die Professionalität des Handelns zu beurteilen, ist es wichtig, dies begleitend zu machen. Dafür muss schon zu Beginn einer Maßnahme gemeinsam geklärt werden, was erreicht werden soll und wie das überprüft werden kann. Eingangsmessungen und eine bewusste Auswahl der Datengrundlagen sind dafür unerlässlich. Wenn dieser Prozess früh und strukturiert angegangen wird, ist Wirkungsmessung gut machbar.

Wie lässt sich die Weiterbildung als Forschungsfeld beschreiben?

MARTIN: Die Weiterbildung ist ein extrem heterogenes Feld, was die Wirkungsmessung besonders anspruchsvoll macht. Die erzielten Messungen sind dann auch immer domänenspezifisch: Ein Gesundheitskurs, der etwa auf gesündere Ernährung abzielt, entfaltet ganz andere Wirkungen als ein Kurs der politischen Bildung oder ein Sprachkurs für Zugewanderte. In der Praxis bedeutet das: Die Indikatoren und Methoden müssen immer zum jeweiligen Thema und zur Zielgruppe passen. Man

muss sich also intensiv mit der jeweiligen Ausrichtung der Maßnahme und ihren Outcomes auseinandersetzen. Zwar gibt es auch theoretische Annahmen dazu, dass Weiterbildung an sich einen Effekt hat, beispielsweise, dass kognitive Skills oder allgemeine Fähigkeiten durch eine Teilnahme verbessert bzw. verbreitert werden. Aber wenn man die konkrete Wirkung einer einzelnen Weiterbildungsmaßnahme untersuchen möchte, dann ist das in der Regel begrenzt auf das, was vermittelt wird.

»Fast alles lässt sich operationalisieren, wenn man sich auf gemeinsame Konventionen und Fragestellungen einigt.«

Andreas Martin

Warum ist die Kausalität so schwer nachzuweisen?

MARTIN: Kausalität ist deshalb so schwer nachzuweisen, weil wir kein kontrafaktisches Universum haben. Wir können immer nur beobachten, wie sich eine Person unter einer bestimmten Bedingung – etwa nach Teilnahme an einer Weiterbildungsmaßnahme – entwickelt. Wir können aber nie gleichzeitig sehen, wie dieselbe Person sich ohne diese Maßnahme entwickelt hätte. Genau diese Information bräuchten wir, um einen »reinen« kausalen Effekt eindeutig zu bestimmen. Das ist das grundlegende Problem: Wir können nie beides – Treatment und Kontrolle – für dieselbe Person gleichzeitig beobachten.

Dennoch gibt es mittlerweile sehr ausgereifte statistische Methoden, mit denen sich diese Herausforderung zumindest teilweise kontrollieren lässt. Durch Randomisierung, Quasi-Experimente oder Längsschnittstudien kann man den kausalen Effekt einer Maßnahme heute relativ verlässlich schätzen und auch die Unsicherheiten dabei abschätzen. Das eigentliche Messen ist also weniger das Problem.

SABINE SCHWARZ: Im Kompetenzzentrum GrundbildungsPFADE begleiten wir aktuell zehn Grundbildungsprojekte. Dabei zeigt sich: Output zu zählen – also etwa, wie viele Teilnehmende einen Kurs besuchen oder wie viele Stunden Unterricht angeboten werden – ist relativ einfach. Aber bei den eigentlichen Outcomes, also den Wirkungen, wird es schnell schwierig, weil es keine Kontrollgruppen oder Längsschnittdaten gibt. Gerade in der Arbeit mit vulnerablen Zielgruppen, etwa jungen Erwachsenen ohne Schulabschluss, ist es oft unmöglich, die Entwicklung über Jahre hinweg systematisch zu verfolgen. Trotzdem sehen wir, dass die intensive Auseinandersetzung mit Wirkungszielen, Zielgruppen und Bedarfen einen echten Mehrwert bringt. Die Mitarbeitenden in den Projekten lernen, genauer zu beschreiben, was sie mit ihrer Arbeit für wen erreichen möchten und wie sie diese angestrebten Veränderungen analysieren können. Das ist zwar aufwendig, hilft aber dabei, den Wert der eigenen Arbeit und die Effekte und somit auch den Erfolg besser sichtbar zu machen.

Wie relevant sind Forschungsergebnisse für die Praxis? Großangelegte Studien haben oft eine große »Flughöhe«.

SCHWARZ: Das ist tatsächlich ein Thema, das uns in der Praxis immer wieder begegnet. Gerade bei großangelegten Studien – die oft diese gewisse »Flughöhe« haben – ist das Übertragen der

Ergebnisse auf die konkrete Arbeit vor Ort schwierig. So zeigt etwa die Studie vom Deutschen Zentrum für Integrations- und Migrationsforschung, dass Integrationskurse einen deutlichen Einfluss auf die Beschäftigungsquote Geflüchteter haben. Für Lehrende in Integrationskursen ist das zwar erfreulich, aber für die alltägliche Praxis bleibt der Nutzen oft abstrakt. Viele Lehrende beobachten Wirkungen bei ihren Teilnehmenden – wie aktivere Teilhabe, mehr Selbstvertrauen oder bessere Kommunikationsfähigkeit. Dies ist für Handeln von Lehrenden oft viel relevanter als eine Zahl, die Auskunft über den Zusammenhang von Integrationskursen und Arbeitsmarktintegration gibt. Misstrauen in der Bildungspraxis besteht zudem hinsichtlich des Einflusses von großangelegten Studien auf politische Entscheidungsprozesse. Bleiben wir beim Beispiel Integrationskurse: Wenn trotz nachgewiesener positiver Wirkungen seitens der Politik Mittel gekürzt werden, stellt die Bildungspraxis in einer eher fatalistischen Perspektive durchaus auch den generellen Nutzen solcher großangelegten Studien in Frage. Ein anderes Thema, was ich im Bereich der Grundbildungspraxis wahrnehme, ist, dass die Erwartungen an die Effekte der Bildungsangebote oft groß sind. Menschen sollen »danach« etwa ihren Alltag besser gestalten können oder kompetenter mit Desinformation umgehen können. Hier stellt sich durchaus die Frage ob solche Erwartungen auch überzogen sind. Als drittes beobachte ich, dass viele das Thema Wirkungsmessung als bedrohlich erleben, weil sie das Gefühl haben, sich ständig rechtfertigen müssen, ob ihre Arbeit »wirkt«.

Ist die Unterscheidung der Wirkung auf Makro-, Meso- und Mikroebene entscheidend?

MARTIN: Absolut. Viele Studien arbeiten zwar auf der Makroebene und betrachten gesellschaftliche Entwick-

lungen, doch die eigentlichen Wirkmechanismen laufen meist über die Mikroebene: Politische Impulse oder Förderprogramme entfalten ihre Wirkung erst, wenn einzelne Menschen daran teilnehmen und sich weiterentwickeln. Diese individuellen Veränderungen summieren sich über Organisationen und Lernorte hinweg zu Effekten auf gesellschaftlicher Ebene. Auf diesem Weg können jedoch unerwartete

»Viele empfinden Wirkungsmessung als bedrohlich, weil sie befürchten, sich rechtfertigen zu müssen, ob ihre Arbeit ›wirkt‹.«

Sabine Schwarz

Folgen auftreten, sodass am Ende weniger – oder andere! – Wirkung erreicht wird, als erhofft. Deshalb ist es wichtig, alle Ebenen im Blick zu behalten und Wirkung nicht nur »oben« oder »unten« zu messen. Wer gesellschaftspolitisch steuern will, muss von Beginn an erfassen, was auf allen Ebenen passiert – und Maßnahmen wissenschaftlich begleiten, um nachvollziehen zu können, warum bestimmte Effekte eintreten oder nicht.

SCHWARZ: Das kann ich nur bestätigen – gerade auf Projektebene ist es sinnvoll, die Wirkungslogik von hinten aufzurollen. Zuerst sollte gemeinsam überlegt werden, welchen Impact und Outcome man eigentlich erreichen will, und dann rückwärts geplant werden: Welche Ressourcen stehen zur Verfü-

gung, wer sind die Zielgruppen, welches Problem soll gelöst werden? So kann man auf den verschiedenen Ebenen immer wieder überprüfen, ob man noch auf dem richtigen Weg ist. Das ist für viele ungewohnt, aber genau dieser Ansatz – vom gewünschten Ergebnis rückwärtszudenken und die nötigen Schritte gezielt zu planen – macht es möglich, Wirkungen nachvollziehbar zu analysieren. Statt erst am Ende zu fragen, ob eine Maßnahme etwas gebracht hat, sollte man den Prozess besser kontinuierlich begleiten und immer wieder reflektieren, ob man noch auf dem richtigen Weg ist.

Wie sieht das Vorgehen bei geförderten Projekten aus – gibt es Unterschiede?

SCHWARZ: Bei geförderten Projekten werden die Ziele meist auf mehreren Ebenen festgelegt: Bund oder Land geben übergeordnete, oft politisch motivierte Vorgaben – so sollen etwa bei der Förderlinie GrundbildungsPFADE Bildungsübergänge in den Blick genommen werden und Bildungslücken identifiziert und geschlossen werden. Ein zentrales Ziel ist es dabei, dass erfolgreiche Ansätze nach dem Auslaufen der Förderung verstetigt werden, zum Beispiel durch die Länder selbst, um so langfristig echte strukturelle Veränderungen zu bewirken. Dies passiert bislang im Projektgeschäft allerdings eher selten. Eine stärker wirkungsorientierte Ausrichtung der Projektarbeit könnte dabei aber durchaus ein paar belastbare Argumente auch in Richtung Verstetigung liefern.

Unterscheidet sich die wissenschaftliche Wirkungsmessung hiervon?

MARTIN: Die eigentliche Herausforderung liegt weniger in den Methoden, sondern darin, wie Wirkungsmessung genutzt wird. In der Praxis besteht häufig die (nicht unberechtigte) Be-

fürchtung, dass solche Evaluationen der Rationalisierung und Steuerung dienen, es also darum geht, mit weniger Ressourcen auszukommen oder Verantwortlichkeiten für nicht erreichte Ergebnisse zuzuweisen. Gerade bei der Überprüfung gesellschaftlicher Ziele, wie bei staatlich geförderten Sprachkursen, darf der Umgang mit den Ergebnissen nicht in eine solche negative Fehlerkultur führen. Ohne systematische Auswertung bleiben Lernerfahrungen aus, und Projekte verschenken wichtige Chancen zur Verbesserung.



PROF. DR. ANDREAS MARTIN

ist Leiter der Abteilung System und Politik am Deutschen Institut für Erwachsenenbildung (DIE) und Professor für Bildungswissenschaft unter besonderer Berücksichtigung des Systems der Weiterbildung und seiner Adressatinnen und Adressaten an der FernUniversität in Hagen.

martin@die-bonn.de

SCHWARZ: Wenn das Ziel – wie etwa die Arbeitsmarktintegration – klar vorgegeben ist, ist auch klar, was und wie das gemessen werden kann. Oft sind die Ziele aber nicht so eindeutig zu quantifizieren. Wenn es zum Beispiel um Teilhabe geht – was ist genau damit gemeint? In Projekten erleben wir, dass sie sich viel stärker mit Wirkungsanalyse identifizieren, wenn sie selbst an der Entwicklung und Formulierung der Wirkziele beteiligt sind. Sobald sie mitbestimmen können, was gemessen werden soll und wie, empfinden sie die Evaluation nicht mehr als von außen aufgezwungen, sondern als nachvollziehbaren und sinnvollen Prozess. Das fördert die Bereitschaft,

sich auf Evaluation einzulassen, diese mitzugestalten und gemeinsam zu lernen. In der Bildungspraxis ist dieses Vorgehen jedoch selten. Hier überwiegt oft die Angst, durch Wirkungsevaluationen Nachteile zu erleiden oder Fördermittel zu verlieren. Viele Akteure möchten sich vielleicht auch nicht »in die Karten schauen« lassen, was ressortübergreifende Zusammenarbeit und echte Weiterentwicklung erschwert. Dabei liegt der Fehler allerdings nicht unbedingt bei einzelnen Akteuren selbst, die in eine Verweigerungshaltung gehen, sondern auch im System. Kein Träger vermittelt Teilnehmende weiter, selbst wenn er vielleicht sogar merkt, dass die Person bei ihm in der falschen Maßnahme ist. Er würde sonst die Kopfpause verlieren. Solche »Systemfehler« lassen sich in vielen Bereichen finden. Die Sorge vor Zuständigkeitsverlust blockiert dabei den offenen Umgang mit Wirkung und verhindert, dass gemeinschaftlich an Verbesserungen gearbeitet wird.

Ist das gerechtfertigt?

MARTIN: Bei fast allen Projekten und Maßnahmen ist das Lehrgeld so oder so gezahlt worden – Fehler und Rückschläge gehören zum Entwicklungsprozess dazu. Entscheidend ist, daraus zu lernen und Fehler nicht als Makel zu sehen. In der Wissenschaft ist diese Fehlerkultur selbstverständlich: Fortschritt entsteht durch Versuch und Irrtum, und auch Null-Ergebnisse werden offen publiziert. Genau diese Offenheit fehlt oft in der Praxis der Weiterbildung – dort werden Fehler eher verschwiegen oder verborgen statt diese als Chance zur Weiterentwicklung zu nutzen. Wir zahlen das Lehrgeld, aber wir lernen nicht die teuer erkaufte Lektion. Ich befürchte, auch bei den Fördermittelgebern ist das so: Um wirklich zu schauen, wo bei Fördervorhaben die gesteckten Ziele erreicht wurden, wo es Hindernisse und Barrieren gibt, müsste man

diese Programme von Beginn an wissenschaftlich begleiten, die relevanten Daten erheben oder Interventionen planen. Doch in der Regel kommt die Politik erst am Projektende, wenn überhaupt, und sagt: Jetzt würden wir gerne wissen, ob das was gebracht hat – und dann ist es schwierig bis unmöglich, genaue Antworten zu liefern. Zudem ist es entscheidend, nicht nur herauszufinden, ob ein Projekt oder Intervention wirksam war; sondern es geht um die Heterogenität der Effekte: was wirkt und was nicht, und bei wem wirkt es und bei wem nicht.



DR. SABINE SCHWARZ

ist Leiterin des Bereichs Grundbildung in der Lernenden Region – Netzwerk Köln und Teil des Kompetenzzentrums GrundbildungSPFADE.

sabine.schwarz@bildung.koeln.de

Ein großes Problem der Wirkungsmessung scheint mir auch, dass wir es in Lehr-Lern-Prozessen mit sozialen Konstellationen zu tun haben, bei der viele Faktoren zum Erfolg oder Scheitern beitragen können. Wie gut lassen sich Wechselwirkungen unterschiedlicher Faktoren messen? Ist das nicht sehr aufwendig?

MARTIN: Wechselwirkungen in der Weiterbildung zu messen, ist tatsächlich komplex, weil viele Faktoren sich gegenseitig beeinflussen und verschiedene kausale Richtungen möglich sind. Dennoch gibt es heute ausgefeilte Modelle und Methoden, mit denen sich auch

solche Zusammenhänge analysieren lassen. Moderne Modellierungen ermöglichen es, spezifische Effekte – etwa von X auf Y – herauszuarbeiten, selbst wenn die Wirkmechanismen verschachtelt sind. Natürlich ist das aufwendig und erfordert einen gewissen Ressourceneinsatz. Wichtig ist dabei, dass wir klare Kriterien und ein gutes Verständnis dafür haben, wie belastbar einzelne Ergebnisse sind. Man kann beispielsweise auch die Irrtumswahrscheinlichkeit benennen und transparent machen, wie sicher eine Aussage ist.

Kann man bei der Wirkungsmessung jemals absolute Sicherheit erreichen?

MARTIN: Perfekte Gewissheit gibt es in diesen komplexen Systemen nie. Niemand kann alle Kausalketten kennen. Aber fast alles lässt sich operationalisieren, wenn man sich auf gemeinsame Konventionen und Fragestellungen einigt. Die Lösungen sind oft nicht perfekt, doch sie ermöglichen es, Ergebnisse realistisch einzuschätzen und offen zu kommunizieren, was bekannt ist und wo Unsicherheiten bleiben. So lassen sich auch sehr komplexe Wechselwirkungen grundsätzlich messen und beurteilen, auch wenn immer ein Rest Unsicherheit bleibt.

Oft treten Wirkungen gerade in der Bildung ja zeitlich stark verzögert ein. Gibt es Wirkungen, die man nicht messen kann?

MARTIN: Das hängt stark davon ab, wie konkret die Fragestellung ist. Grundsätzlich lassen sich auch langfristige oder indirekte Effekte von Bildung – etwa auf Persönlichkeits- oder Identitätsentwicklung – erfassen, wenn man sie präzise genug fasst. In Panelstudien kann man beispielsweise Zeitabstände einbauen und prüfen, ob nach einem, zwei oder mehreren Jahren messbare Veränderungen auftreten.

Also auch so etwas Individuelles wie Persönlichkeitsentwicklung lässt sich messen?

MARTIN: Gerade bei Themen wie Identitätsbildung oder Werthaltungen ist es entscheidend, das Untersuchungsinteresse klar zu definieren: Wenn etwa gefragt wird, ob sich zum Beispiel durch Weiterbildung die Identifikation mit dem Herkunfts- oder Aufnahmeland verändert – was ich als Teil der Persönlichkeitsstruktur fassen würde –, lässt sich das mit geeigneten Daten und Instrumenten untersuchen. Zu allgemein formulierte Fragen – wie die Wirkung von Bildung auf »Identität« insgesamt – sind dagegen zu global, um sie empirisch zuverlässig zu messen.

SCHWARZ: Ich bin auch der Meinung, dass so individuelle Veränderungen wie Persönlichkeitsentwicklung sich messen lassen – allerdings nur, wenn man zuvor sehr genau klärt, was damit gemeint ist. Solche Begriffe sind Container-Begriffe, wie auch der Begriff »Teilhabe«. Wenn wir sagen: »Bildung verbessert Teilhabechancen«, dann muss man Teilhabe konkretisieren, um sie messbar zu machen: Wie zeigt sich Teilhabe im Alltag? Welche Facetten gehören dazu? Was tun Menschen, die teilhaben, anders als jene, die es nicht tun?

Also ist Wirkungsmessung am Ende auch immer eine Frage der Definition?

SCHWARZ: Tatsächlich. Erst durch die sogenannte Operationalisierung – also das Übersetzen abstrakter Begriffe in beobachtbare Merkmale – könnten auch Veränderungen im Bereich der Persönlichkeitsentwicklung oder einer Teilhabeerweiterung analysiert werden. Dabei ist wichtig, sowohl individuelle Aspekte als auch systemische Faktoren, wie Barrieren oder Ausschlussmechanismen zu berücksichtigen. Anders als bei Wissen oder Kompetenzen, die sich mit Tests vor und nach einer Maßnahme recht eindeutig erfassen lassen, ist das bei komplexen

Begriffen wie »Teilhabe« deutlich schwieriger. Hier sind präzise Definitionen und eine enge Zusammenarbeit von Wissenschaft und Praxis unerlässlich, um sinnvolle Messungen zu ermöglichen.

Was möchten sie abschließend betonen?

MARTIN: Die mangelnde Fehlerkultur ist in meinen Augen das größte Hindernis bei der Wirkungsmessung: Solange wir nicht bereit sind, auch Misserfolge offen anzusprechen und daraus zu lernen, werden Projekte und Programme nicht systematisch besser. Das betrifft alle Ebenen – von der Praxis bis zur Politik. Niemand lässt sich gern darauf festnageln, dass eingesetzte Mittel vielleicht nicht zum gewünschten Erfolg geführt haben. Aber genau diese Offenheit und Bereitschaft zur Evaluation wäre notwendig, um wirklich Fortschritte zu erzielen und Bildungsangebote nachhaltig weiterzuentwickeln.

SCHWARZ: Ich wünsche mir, dass Bildungseinrichtungen selbstbewusster mit dem Thema Wirkung umgehen und sich trauen, Bildungsziele wie Selbstbewusstsein, Lebensqualität oder Teilhabe aktiv mitzuformulieren. Gerade die vielen Erfolgsgeschichten aus dem Alltag zeigen, wie wertvoll auch subjektiv erlebte Wirkungen sind. Diese sollten von der Bildungspraxis selbstbewusst benannt und in die Diskussion eingebracht werden. Ein Beispiel: Eine Teilnehmerin berichtete, dass sie nach einem Grundbildungskurs erstmals selbstständig ein Gespräch bei der Bank führt. Für sie war es ein riesiger Schritt, der in keiner Statistik auftaucht, aber ihre Lebensqualität spürbar verbessert hat. Solche »gefühlten« Erfolge sind schwer zu messen, aber sie sind echte Wirkung und sollten von der Bildungspraxis selbstbewusst benannt und in die Diskussion eingebracht werden.

Ich danke Ihnen sehr für dieses Gespräch!