

MINT-Projekte für Schülerinnen an Hochschulen. Analyse des Wirkungsmechanismus und Meta- Evaluation der empirischen Evidenz

Stefanie Brenning, Elke Wolf

Zusammenfassung: An deutschen Hochschulen für angewandte Wissenschaften existieren zahlreiche Angebote, um Schülerinnen für Studiengänge und Berufe im MINT-Bereich zu begeistern. Wissenschaftlich gesicherte Erkenntnisse über deren Wirksamkeit fehlen jedoch weitgehend. Ausgehend vom aktuellen Forschungsstand entwickeln wir eine mehrstufige Wirkungslogik dieser Maßnahmen und analysieren anschließend die Methodik der uns vorliegenden 13 Evaluationen mithilfe einer Meta-Evaluation. Die Befunde zeigen, dass diese Evaluationen von MINT-Projekten nur sehr eingeschränkt der Erfolgsmessung dienen, da sich die schriftlichen Fragebögen größtenteils auf kurzfristige Erfolgsindikatoren beschränken und die Untersuchungsdesigns nicht den Ansprüchen einer Kausalanalyse genügen.

Schlüsselwörter: Wirkungslogik, MINT-Projekte, Meta-Evaluation, Programmevaluation, Gleichstellung, Hochschule

STEM projects for female pupils at Universities of Applied Sciences. Analysis of impact mechanisms and meta-evaluation of empirical evidence.

Abstract: Universities of Applied Sciences in Germany offer numerous programs for female pupils in order to spark their interest in studying and working in STEM disciplines. Yet, scientifically verified impact assessments of these programs are still largely missing. Proceeding from the current state of research, we develop a multi-level impact logic of these programs and subsequently analyze the applied methods in 13 evaluations that are accessible to us using a meta-evaluation. Our results show that these evaluations of STEM projects are constrained in their ability to measure success because their questionnaires mostly focus on very short-term success indicators and their research designs don't fulfill the requirements of causal analyses.

Keywords: impact logic, STEM projects, meta-evaluation, program evaluation, gender equality, higher education

1 Einleitung

Der Anteil weiblicher Studierender an Hochschulen und Universitäten und weiblicher Fachkräfte im MINT¹-Bereich liegt nach wie vor weit hinter dem Anteil ihrer männlichen Mitstreiter. Während der Frauenanteil in MINT-Studiengängen zum Wintersemester 2018/19 insgesamt knapp ein Drittel beträgt (30.9%), liegt dieser in den Ingenieurwissenschaften und der Informatik nur etwas über einem Fünftel (23.6% bzw. 21.4%) (Statistisches Bundesamt, 2019). Der Frauenanteil unter den erwerbstätigen MINT-Akademikerinnen und -Akademikern betrug im Jahr 2015 21.5%, zugleich war der Anteil erwerbstätiger weiblicher MINT-Fachkräfte an allen erwerbstätigen MINT-Fachkräften mit 11% noch deutlich niedriger (Anger, Koppel & Plünnecke, 2018, S. 14f.). Dies steht in Kontrast zu den im April 2018 486,600 offenen Stellen im MINT-Bereich, die nicht besetzt werden konnten (Anger et al., 2018, S. 50).

Die Gründe der Unterrepräsentanz von Frauen im MINT-Bereich sind vielfältig. Neben dem Hindernis, als eine von wenigen Frauen in einem männerdominierten Umfeld zu studieren und/oder zu arbeiten, zeigt sich auch, dass geschlechterstereotype Berufsbilder, falsche Vorstellungen von MINT-Berufen und eine geringere Selbsteinschätzung junge Frauen von der Wahl eines MINT-Berufes abhalten (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung [BLK], 2002; Ertl, Luttenberger & Paechter, 2014; Pöllmann-Heller & Rudolph, 2019; Wentzel, 2008; Weinhardt, 2017).

Die Bundesregierung initiiert und unterstützt seit einigen Jahren zahlreiche Programme, um Frauen für MINT-Berufe, insbesondere im Bereich der Informationstechnologie, zu gewinnen (z.B. Komm, mach MINT; Girls' Day; MINT-E-Plattform; Initiative Klischeefrei).

An Hochschulen und Universitäten liegt der Schwerpunkt der Aktivitäten zur Gewinnung von Frauen für ingenieur- und naturwissenschaftliche Studiengänge im Bereich der Kontext- und Begleitmaßnahmen für Schülerinnen (Löther & Girlich, 2011). Hierzu zählen in erster Linie Informationsveranstaltungen und Workshops zur Forschung und Anwendung im MINT-Bereich. Ziel dieser Angebote ist es in der Regel, das Interesse an ingenieur- und naturwissenschaftlichen Fragestellungen zu erhöhen, die Studien- und Berufsfindung mit Informationsangeboten zu unterstützen und die Einschätzung der eigenen Fähigkeiten zu verbessern. Angesichts der Fülle der Maßnahmen und den nicht unerheblichen Ressourcen, die hierfür investiert werden, stellt sich dringend die Frage nach den Wirkungen dieser Maßnahmen.

Aber wann ist ein MINT-Projekt erfolgreich? Was kann bewirkt werden und wie kann dies gemessen werden? Auch wenn die Maßnahmen darauf abzielen, den Frauenanteil in den MINT-Berufen zu erhöhen, wird die alleinige Teilnahme an einem MINT-Projekt die Studien- und Berufswahlentscheidung nicht monokausal beeinflussen. Die Studien- wie auch die Berufswahl ist komplex und von vielen Faktoren abhängig, auf die spezielle Angebote für Mädchen keinen Einfluss nehmen können. Dies gilt beispielsweise für die frühe geschlechtsspezifische Sozialisation, die vorwiegend innerhalb der Familie und durch die Eltern stattfindet. Gleichwohl ist Sozialisation ein Lernprozess, der nicht im Kindesalter abgeschlossen ist, sondern die kontinuierliche Auseinandersetzung mit der sozialen Umwelt

1 MINT bezeichnet die Fächergruppe der Mathematik, Informatik, Naturwissenschaften und Technik.

umfasst (Kray, 2019; Siegler, Eisenberg, De Loache & Saffran, 2016). In späteren Lebensphasen können dann unterschiedliche Personen und Institutionen, wie die Hochschule und damit verbundene Fördermaßnahmen, zur Sozialisation beitragen (Hurrelmann, Bauer, Grundmann & Walper, 2015). Demnach kann erwartet werden, dass die Teilnahme an einem MINT-Projekt einer Hochschule für Angewandte Wissenschaften für Schülerinnen die veränderbaren Determinanten der Studien- und Berufswahl positiv beeinflusst und relevante Hürden bei der Entscheidung für den MINT-Bereich reduziert. Hierzu zählen u.a. die Selbsteinschätzung ihrer Fähigkeiten sowie das Wissen über verschiedene MINT-Berufe. Die Wirksamkeit einer Maßnahme lässt sich demnach anhand jener empirisch relevanten Determinanten der Studien- und Berufswahl messen, welche durch die Teilnahme an einer Maßnahme verändert werden können.

Löther und Girlich (2011) zeigen allerdings in ihrer Meta-Analyse u.a., dass in Evaluationen von MINT-Schülerinnen-Projekten die Zufriedenheit der Teilnehmerinnen meistens als das zentrale Erfolgskriterium verwendet wird. Nur wenige Evaluationsstudien zu Fördermaßnahmen im MINT-Bereich für Mädchen messen den Erfolg anhand von Einstellungsänderungen (vgl. u.a. Stöger, Duan, Schirner, Greindl & Ziegler, 2013; Mokhonko, Nickolaus & Windaus, 2014). Darüber hinaus beschränken sich die meisten Studien auf eine einmalige Befragung am Ende der Projekte. Belastbare empirische Befunde zur Effektivität der MINT-Schülerinnen-Projekte liefern Stöger et al. (2013), Hiller (2014), Kenkmann (2005) und Zeisberg (2013), die in Prä-Post-Befragungen positive Effekte auf relevante Studien- und Berufswahl-Determinanten nachweisen. So zeigen Stöger et al. (2013) mit ihrer Evaluation eines Online-Mentoring-Programms für Schülerinnen von 11 bis 18 Jahren, dass die Teilnehmerinnen bei sechs von sieben Determinanten Treatment-Effekte aufweisen. Die Schülerinnen verfügten nach Abschluss des Mentoring-Programms im Vergleich zu einer Kontrollgruppe u.a. über ein größeres Wissen über Studiengänge und Berufe in MINT. Aber auch bei der Intention, nach Abschluss der Schule ein MINT-Studium aufzunehmen, einem der längerfristig orientierten Erfolgsindikatoren, zeigten sich positive Effekte. Mokhonko (2016) untersucht die Effekte von Schülerlaboren an Gymnasien für die siebte bis zehnte Klasse mit Durchführungsdauern zwischen einer bis 18 Stunden auf das fachspezifische Fähigkeitsselbstkonzept, die fachspezifischen und beruflichen Interessen sowie die naturwissenschaftlich-technischen Bildungs- und Berufsvorhaben. Insgesamt zeigten sich nur punktuell positive und insgesamt geringe Effekte, z.B. stieg bei Schülerinnen der siebten Klasse das Interesse am intellektuell-forschenden Bereich nach Teilnahme am Schülerlabor signifikant an. Es fällt weiterhin auf, dass sich auch für Maßnahmen längerer Dauer wider Erwarten keine positiven Fördereffekte feststellen lassen konnten. Löther und Girlich (2011, S. 96) kommen zu der Schlussfolgerung, dass die existierenden Evaluationen oftmals der „Entwicklung und Verbesserung des Programms sowie Legitimation innerhalb der Hochschule und gegenüber den Geldgebern“ dienen und nur sehr begrenzt zur objektiven Überprüfung der Wirksamkeit der Projekte und des tatsächlichen Effekts auf die nachfolgende Studien- und Berufswahl der Teilnehmerinnen im MINT-Bereich beitragen.

Insbesondere für die Projekte an Hochschulen für angewandte Wissenschaften (HAW), die aufgrund der hohen Durchlässigkeit beim Studiengang, den Kooperationen mit Unternehmen und dem hohen MINT-Anteil besonders gute Bedingungen zur Gewinnung von MINT-Fachkräften bieten, steht eine Analyse der Evaluationen von Schülerinnen-Projekten

zur Gewinnung von Frauen im MINT-Bereich noch aus. Wir konzentrieren uns daher auf MINT-Fördermaßnahmen für Schülerinnen, die an deutschen HAW durchgeführt werden.²

Ziel dieser Studie ist es, existierende Evaluationen von MINT-Schülerinnen-Projekten anhand ausgewählter Indikatoren zu analysieren. Wir gehen in mehreren Schritten vor. Zu Beginn beschreiben wir den Wirkungsmechanismus der Maßnahmen anhand eines Logic Charts und identifizieren verschiedene Indikatoren auf den drei Wirkungsebenen Output, Outcome und Impact. Diese theoretische und empirische Herleitung der intendierten Effekte kann zukünftig auch zur Konzeption von Projekten und deren Evaluationen genutzt werden. Uns dient die Wirkungslogik im Folgenden der Klassifizierung der in den Evaluationen gemessenen Erfolgsindikatoren. Der empirische Schwerpunkt unserer Untersuchung ist in Abschnitt 3 dargestellt: Wir analysieren die Methodik der vorliegenden Evaluationen anhand der existierenden Standards der Evaluationsforschung, die sich auf Basis der vorliegenden Evaluationsberichte untersuchen lassen. Hierzu zählen die Auswahl der Erfolgsindikatoren sowie die Validität der Wirkungsanalyse. Somit können die Zuverlässigkeit der Erkenntnisse bewertet und Verbesserungspotenziale für die Planung und Durchführung zukünftiger Evaluierungen aufgezeigt werden (Scriven, 1969; Widmer, 1996; Patton, 1997).

Von einer inhaltlichen (Meta-)Analyse wird aufgrund der eingeschränkten Vergleichbarkeit der Erfolgsindikatoren und den mediokren Befunden der Meta-Evaluation abgesehen. Die eingeschränkte Aussagekraft und geringe Anzahl an Evaluationen verdeutlichen jedoch unweigerlich die Dringlichkeit, die Wirkungsmessung von MINT-Projekten zu thematisieren und zu professionalisieren.

2 Die Wirkungslogik von MINT-Projekten für Schülerinnen

Bei der Herleitung einer Wirkungslogik konzentrieren wir uns auf Maßnahmen für Schülerinnen, die deren MINT-Interesse in erster Linie durch eigene Aktivitäten wecken wollen und ihnen Einblicke in das Studium an HAW und angewandte Forschungsfragen gewähren, da der potentielle Wirkungsmechanismus dieser Projekte sehr ähnlich ist.

Auf Basis einer Umfrage an allen HAW zeigen Brenning und Wolf (2019), dass die Angebote zur Studien- und Berufsorientierung für Schülerinnen sehr vielfältig und unterschiedlich sind. Dennoch verfolgen alle MINT-Projekte für Schülerinnen die Ziele, die Teilnehmerinnen mithilfe von ein- oder mehrtägigen Veranstaltungen über naturwissenschaftlich-technische Fragestellungen, Themen und Studiengänge zu informieren, sie an MINT-Fragestellungen heranzuführen und letztendlich für eine Studien- und Berufswahl im MINT-Bereich zu begeistern. Die Schülerinnen nehmen dafür an Workshops, Vorlesungen und Informationsveranstaltungen teil oder führen Gespräche mit Studierenden, Lehrenden,

2 Das dieser Analyse zugrundeliegende Verbundvorhaben „MINT-Strategien 4.0 – Strategien zur Gewinnung von Frauen für MINT-Studiengänge an Hochschulen für angewandte Wissenschaften“ wird mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01FP1714 und 01FP1715 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen. Wir danken Andrea Löther (GESIS – Leibniz-Institut für Sozialwissenschaften, Köln) für zahlreiche und konstruktive Diskussionen sowie zwei anonymen Gutachter/innen für wertvolle Kommentare zu unserem Manuskript.

Hochschulmitarbeiterinnen und Hochschulmitarbeitern oder Frauen, die im MINT-Bereich arbeiten. Zumeist besichtigen sie Hochschulräumlichkeiten wie Labore und Hörsäle und lernen Studiengänge aus dem MINT-Bereich kennen. Teils bearbeiten die Schülerinnen konkrete Fragestellungen, experimentieren mit naturwissenschaftlich-technischen Phänomenen und machen sich mit der angewandten Forschung an HAW vertraut. Viele Fördermaßnahmen an HAW kooperieren auch mit Unternehmen aus ihrer Region, indem die Schülerinnen Unternehmen besuchen und Einblicke in MINT-Berufsbilder und den Arbeitsalltag, z.B. einer Ingenieurin, erhalten können (Brenning & Wolf, 2019).

Die Wirkungslogik der Schülerinnen-Projekte beschreiben wir mithilfe eines Logic Charts, dessen Wirkungsebenen im darauffolgenden Kapitel zur Untersuchung der Evaluationsberichte dienen. Das logische Modell wurde von der W. K. Kellogg Foundation als Evaluierungswerkzeug entwickelt, um Maßnahmen effektiv zu planen, zu implementieren und zu evaluieren. Ein Logic Chart visualisiert die Beziehungen zwischen Aktivitäten, Programminhalten sowie kurz-, mittel- und langfristigen Wirkungen (W. K. Kellogg Foundation, 2004, S. III). Unsere Wirkungsanalyse der Fördermaßnahmen für Schülerinnen baut auf dem Logic Chart von Löther und Girlich (2011) auf, welches das allgemeine logische Modell auf die Funktions- und Wirkungsweisen von Maßnahmen zur Studien- und Berufsorientierung für Schülerinnen anwendet.³

Die Wirkungen einer Maßnahme lassen sich typischerweise in drei Ebenen einteilen: Output, Outcome und Impact (Döring & Bortz, 2016). Für eine umfassende Bewertung einer Maßnahme sollten Evaluationen die Wirkungen auf allen drei Ebenen des Logic Charts (s. Abbildung 1) erfassen.

Der Output, die unmittelbaren Ergebnisse einer Maßnahme, umfasst im Allgemeinen Indikatoren der tatsächlich erbrachten Leistungen und Angebote einer Intervention, der Inanspruchnahme der Leistungen sowie die Zufriedenheit mit den Angeboten. In unserem Logic Chart stellt der Output (1.) die Akzeptanz der Maßnahme bei den Teilnehmerinnen und somit eine Grundvoraussetzung für den Erfolg einer Maßnahme dar. Dies zeigt sich daran, ob die angebotenen Plätze ausgebucht sind, wie auch in der Bewertung der Veranstaltungen, der Inhalte und der gesamten Fördermaßnahme (1.1), in der Bewertung der beteiligten Personen, z.B. der Organisatorinnen und Organisatoren sowie der Referentinnen und Referenten (1.2), und anhand einer potenziellen Weiterempfehlung des Schülerinnen-Projekts wie auch der Frage, ob die Teilnehmerinnen selbst wieder an einem MINT-Projekt für Schülerinnen teilnehmen würden (1.3).

Die zweite Stufe des logischen Modells, der Outcome (2.), beschreibt alle kurzfristigen Wirkungen, die vom Output bei der Zielgruppe verursacht werden (Döring & Bortz, 2016). Wir berücksichtigen hier jene empirisch relevanten Determinanten der Studien- und Berufswahl, welche durch die Teilnahme an einer Maßnahme verändert werden können. Hierzu zählen die potenzielle Steigerung des Interesses der Schülerinnen am MINT-Bereich (2.1), die Unterstützung bei der Studien- und Berufsfindung (2.2), die Erhöhung der Selbstwirksamkeit und Selbsteinschätzung (2.3) sowie die Reduktion von Geschlechterste-

3 Wir begrenzen das Logic Chart jedoch auf die geplanten bzw. gewünschten Ergebnisse und Wirkungen der Maßnahmen (Output, Outcome und Impact), da der Fokus unserer Meta-Evaluation auf der Bewertung der Erfolgsindikatoren der Maßnahmen liegt und eine Kenntnis des Inputs der Programme keine notwendige Voraussetzung zum Verständnis der Meta-Evaluation der Evaluationen darstellt.

reotypen (2.4). Anhand einer Befragung unter den Koordinator*innen von MINT-Schülerinnen-Projekten an HAW (Brenning & Wolf, 2019) zeigt sich, dass sich diese Outcome-Indikatoren vollständig mit den Zielstellungen von MINT-Schülerinnen-Projekten decken und somit sehr gut geeignet sind, die Erfolge dieser Projekte zu beschreiben.

Grundsätzlich ist das Interesse an Fragestellungen und Phänomenen der MINT-Disziplinen (2.1) ein wichtiger Einflussfaktor für die spätere Studien- und Berufswahl (siehe u.a. Lent, Brown & Hackett, 1994; Taskinen, Schütte & Prenzel, 2013; Obergruber, 2018). Mädchen machen jedoch bereits in der Grundschule weniger Erfahrungen mit Technik als Jungen, da sie in ihrer frühen Kindheit im familiären und weiteren sozialen Umfeld seltener an Aktivitäten mit Technik- oder Forschungsbezug herangeführt werden. Sie entwickeln in der Folge weniger Interesse am technischen Bereich (Solga & Pfahl, 2009). Taskinen, Schütte und Prenzel (2013) machen deutlich, dass u.a. das individuelle Interesse an Naturwissenschaften Einfluss auf das Interesse von Schülerinnen an naturwissenschaftlichen Berufen hat.

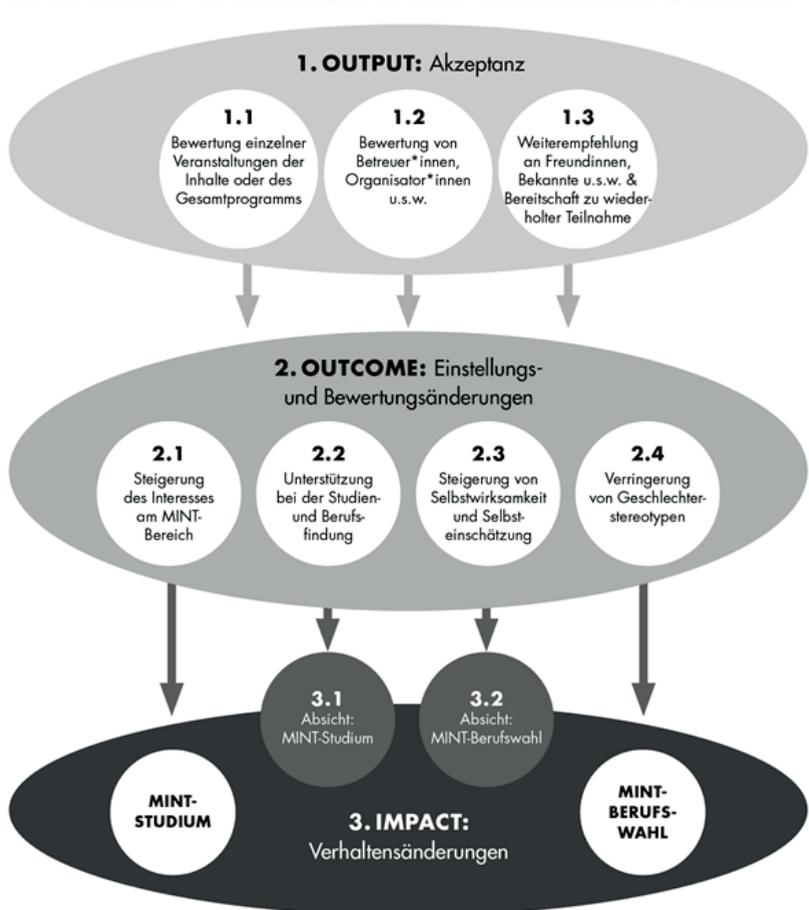


Abbildung 1: Logic Chart – Darstellung der Wirkungsweisen von Maßnahmen zur Studien- und Berufsorientierung für Schülerinnen

Wissenslücken über mögliche Berufsbilder im MINT-Bereich stellen ebenfalls eine relevante Hürde bei der Entscheidung für einen MINT-Beruf dar (siehe hierzu die umfangreiche Darstellung in Matthes, 2019). Aufgrund einer geschlechtsspezifischen Sozialisation kann davon ausgegangen werden, dass Mädchen im Durchschnitt seltener Kontakt zu naturwissenschaftlich-technischen Themenbereichen haben (Taskinen, 2010; Jones, Howe & Rua, 2000; Prenzel, Lankes & Minsel, 2000). Informationen über Studiengänge, Berufe und Arbeitgeber im MINT-Bereich unterstützen junge Frauen bei der Studien- und Berufsfindung (2.2). Diese Informationen können jedoch auch nicht-intendierte Effekte für die Studien- und Berufswahl im MINT-Bereich haben, da sich einige Schülerinnen aufgrund der neu gewonnenen Informationen auch explizit gegen einen MINT-Beruf entscheiden könnten.

Eine wichtige Rolle spielen ebenso die Selbsteinschätzung und die Selbstwirksamkeit der Schülerinnen (2.3): Die Ergebnisse von Weinhardt (2017) zeigen, dass sich Schülerinnen im Vergleich zu Schülern im Fach Mathematik selbst bei gleichen Noten schlechter einschätzen. Darüber hinaus zeigt Correll (2001, 2004), dass Mädchen ihre Erfolge in Mathematik und Naturwissenschaften eher auf Glück und Anstrengung zurückführen, während Jungen den Erfolg als Konsequenz ihrer Kompetenz betrachten. Mädchen entscheiden sich demnach tendenziell eher für ein naturwissenschaftliches Fach, wenn sie ihre mathematischen Fähigkeiten positiv einschätzen.

Auch die Selbstwirksamkeit, nach Bandura (1997) die Überzeugung, bestimmte Handlungen erfolgreich ausführen zu können, hat Effekte für die Studien- und Berufswahlentscheidungen von Schülerinnen. Wenn diese grundlegende Überzeugung, z.B. den Anforderungen eines (MINT-)Studiums gewachsen zu sein, bei Schülerinnen in geringerem Maße vorhanden ist, da sie weniger auf ihre Kompetenzen vertrauen (Correll, 2001, 2004), könnte dies nach Banduras Theorie die Wahrscheinlichkeit einer Studien- und Berufswahl im MINT-Bereich senken. Empirische Evidenz dieses Zusammenhangs findet sich u.a. bei Bandura, Barbaranelli, Caprara und Pastorelli (2001) sowie Sheu et al. (2010).

Des Weiteren haben Geschlechterstereotypen (2.4) einen Effekt auf die spätere Studien- und Berufswahl: Zum einen können negative Stereotype hinsichtlich der mathematischen und analytischen Fähigkeiten von Frauen die entsprechenden Leistungen von Schülerinnen mindern, wie es in der Literatur zur Stereotypenbedrohung beschrieben wird (s. u.a. Spencer, Steele & Quinn, 1999). Als Folge davon kann wiederum das Interesse von Mädchen und Frauen am MINT-Bereich sinken (Shapiro & Williams, 2012) und zu einer schlechteren Einschätzung der eigenen Fähigkeiten führen. Zum anderen werden viele MINT-Disziplinen mit Männlichkeit assoziiert und Kompetenzen im MINT-Bereich bei Frauen als unattraktiv und unweiblich wahrgenommen (Ertl et al., 2014; Kessels & Hannover, 2006). In Anlehnung an die Theorie der kognitiven Dissonanz von Festinger (1957) strebt jedes Individuum nach Konsistenz der eigenen Meinungen und Einstellungen und versucht Dissonanzen zwischen nicht miteinander vereinbaren Kognitionen, z.B. eigenen und konträren Ansichten, zu reduzieren. Wenn zwischen dem eigenen geschlechterspezifischen Selbstkonzept und den Stereotypen über Mädchen und Frauen im MINT-Bereich Diskrepanzen vorliegen, distanzieren sich Frauen tendenziell von diesem Berufsbild und wählen daher eher „mädchentypische“ Studien- und Berufsfelder, die nicht im MINT-Bereich liegen (Kessels, 2015), um somit die als unangenehm erlebte kognitive Dissonanz zu reduzieren.

Veränderungen dieser Studien- und Berufswahldeterminanten können unmittelbar nach der Teilnahme an einem Programm gemessen werden. Kurzfristig ist ein Programm also erfolgreich, wenn die beschriebenen Indikatoren auf der Outcome-Ebene durch die Teilnahme an dem Programm positiv beeinflusst werden. Da die Hürden für die Aufnahme eines MINT-Studiums individuell sehr unterschiedlich sind, ist davon auszugehen, dass ein Projekt umso erfolgreicher ist, je mehr positive Teileffekte auf die unterschiedlichen Indikatoren erzielt werden, da somit sehr unterschiedliche Barrieren überwunden werden können. Auch wenn sich aus der Messung der Einstellungen keine direkte Kausalität zur späteren Studien- und Berufswahl herleiten lässt, kann eine in gewünschter Richtung erfolgte Veränderung einer oder mehrerer relevanter Determinanten durchaus die Wahrscheinlichkeit erhöhen, ein MINT-Studium aufzunehmen oder einen MINT-Beruf zu ergreifen.

Auf der dritten Wirkungsebene unseres Logic Charts (Abbildung 1) misst der Impact die langfristigen Wirkungen (auf gesellschaftlicher Ebene) (Döring & Bortz, 2016), die durch verändertes Verhalten oder eine Veränderung der privaten oder beruflichen Situation der Teilnehmerinnen erzeugt werden. Der langfristige Erfolg einer Maßnahme zeigt sich anhand der tatsächlichen Verhaltensänderungen, die ein MINT-Projekt zur Folge haben kann.

Ob die erwünschte Wirkung einer Fördermaßnahme für Schülerinnen, den Frauenanteil in naturwissenschaftlich-technischen Studiengängen zu steigern, erreicht wird, lässt sich idealerweise mithilfe einer Untersuchung der späteren Studien- und Berufswahl der Teilnehmerinnen im MINT-Bereich ermitteln. Die methodische Umsetzung ist jedoch in der Praxis aufgrund fehlender Ressourcen und der Schwierigkeit, die Schülerinnen nach mehreren Jahren zu befragen, kaum realisierbar. In Anbetracht der vielfältigen Einflussfaktoren der Studien- und Berufsentscheidungen wäre damit jedoch noch nicht geklärt, wie relevant die Teilnahme an einem bestimmten MINT-Projekt tatsächlich für diese Entscheidungen war.

Alternativ kann der Impact der MINT-Schülerinnen-Projekte anhand der Intention, ein MINT-Studium (3.1) oder einen MINT-Beruf (3.2) aufzunehmen, gemessen werden. Die Theorie des geplanten Verhaltens von Ajzen (1991) beschreibt ein Modell, welches das Verhältnis zwischen einer Verhaltensintention und dem tatsächlichen Verhalten unter spezifischen Bedingungen erklärt. Ajzen (1991) argumentiert, dass die Verhaltensintention wiederum von Einstellungen bestimmt wird, die durch Erfahrungen und äußere Faktoren beeinflusst werden. Das bedeutet, Verhaltensintentionen sind indirekt durch exogene Ereignisse veränderbar und stehen in direktem Zusammenhang zum tatsächlichen Verhalten. Kim und Hunter (1993) zeigen in ihrer Meta-Evaluation, dass Einstellungen 50% der Varianz bei den bestimmten Verhaltensintentionen erklären und diese wiederum 30% der Varianz im tatsächlichen Verhalten erklären. Auch Armitage und Conner (2001) kommen im Rahmen ihrer Meta-Evaluation von 185 verschiedenen Studien zu dem Ergebnis, dass Intentionen und selbstberichtete Vorhersagen gute Prädiktoren von Verhalten darstellen. Auch wenn die Theorie des geplanten Verhaltens darauf abzielt, den kurzfristigen Zusammenhang zwischen Intention und Handlung zu erklären, zeigen Randall und Wolff (1994), dass der starke Korrelationskoeffizient auch bis zu einer zeitlichen Distanz von 15 Jahren erhalten bleibt. Stöger et al. (2013) nutzen in ihrer Evaluation eines einjährigen Online-Mentoring-Programms für elf- bis 18-jährige Schülerinnen die Studienwahlintentionen einer Versuchs-

und Kontrollgruppe als einen der Erfolgsindikatoren für die Wirksamkeit des Programms. Auch Lent et al. (1994) argumentieren, dass sich die Theorie des geplanten Verhaltens auf die Studien- und Berufswahlentscheidung anwenden lässt, d.h. auch diese Entscheidung wird durch die vorherige Verhaltensintention beeinflusst. Eine in gewünschter Richtung erfolgte Veränderung der Studien- und Berufswahlintentionen erhöht somit die Wahrscheinlichkeit einer bestimmten Studien- und Berufswahl und kann folglich als Erfolg der MINT-Fördermaßnahme interpretiert werden.

Zur Messung der kurzfristigen Effekte von MINT-Projekten können somit die beeinflussbaren Determinanten der Studien- und Berufswahl (Outcome-Indikatoren 2.1 bis 2.4) sowie die Intention einer bestimmten Bildungsentscheidung herangezogen werden. Langfristig zeigt sich der Erfolg anhand eines veränderten Studien- und Berufswahlverhaltens, wobei nicht von einem monokausalen Zusammenhang zu einem einzelnen MINT-Projekt ausgegangen werden kann. Der Blick auf die verschiedenen Teileffekte begegnet darüber hinaus der Kritik von Müller und Albrecht (2016), dass Evaluationen, die sich nur auf den Impact konzentrieren, keine Erkenntnisse über die Wirkungsweise der untersuchten Maßnahmen generieren.

3 Meta-Evaluation der Evaluationen von MINT-Projekten

Im Folgenden analysieren wir die existierenden Evaluationsstudien von MINT-Projekten für Schülerinnen an HAW hinsichtlich ihrer Methodik. In unsere Meta-Evaluation beziehen wir jegliche Evaluationen von MINT-Projekten ein, die explizit für Mädchen konzipiert sind, einen eindeutigen Bezug zu MINT-Studienfächern und -Berufen aufweisen, mindestens einen halben oder mehrere Tage bis zu einem halben Jahr dauern und in den letzten 15 Jahren an einer HAW durchgeführt wurden. Ausgeschlossen wurden Evaluationsberichte von Projekten, die sowohl männliche als auch weibliche Studierende betrachten und deren Analysen und Ergebnisse nicht nach Geschlechtern differenzieren. Des Weiteren haben wir Angebote exkludiert, die nur aus einer kurzen Informationsveranstaltung bestehen.

Insgesamt liegen uns 13 Evaluationsberichte von zehn Einzelmaßnahmen für Schülerinnen vor.⁴ Sieben dieser Maßnahmen werden von den einzelnen Hochschulen selbst durchgeführt. Drei Projekte sind Verbundprojekte, bei denen sich die Hochschulen an Initiativen von Ministerien und Verbänden beteiligen oder mit anderen Hochschulen kooperieren. Alle Angebote, von denen Evaluationsberichte vorliegen, führen direkt nach Abschluss des Projekts eine schriftliche Befragung der teilnehmenden Schülerinnen durch. Elf der 13 Evaluationen wurden hochschulintern durchgeführt.

Die evaluierten Projekte richten sich an Schülerinnen der siebten Klasse bis zum letzten Schuljahr oder Mädchen im Übergang zum Studium und Beruf. Neun der zehn Einzelmaß-

4 Um auch nicht publizierte Evaluationen zu berücksichtigen, haben wir eine umfassende Recherche zu MINT-Projekten für Schülerinnen an HAW durchgeführt. Anschließend wurden die Projektkoordinatorinnen und Projektkoordinatoren von 98 Schülerinnen-Projekten per E-Mail angeschrieben und Projektwebseiten durchsucht, um Berichte bisher durchgeführter Evaluationen zu erhalten (Brenning & Wolf, 2019). Alle Evaluationsberichte wurden uns entweder per E-Mail von Projektkoordinator*innen zugesendet oder standen auf der Webseite des jeweiligen MINT-Angebots zum Download zur Verfügung.

nahmen kooperieren auch mit Unternehmen, um Schülerinnen Einblicke in bestimmte Berufsfelder zu geben. Nicht alle Projekte äußern sich explizit zu den angestrebten Zielen. Eine Bewertung anhand der von den Projekten selbst gesetzten Ziele wäre demnach nicht umfassend genug. Wenn jedoch konkrete Ziele benannt werden, so sind diese immer in den Outcome-Indikatoren unseres Logic Charts inkludiert. Trotz der großen Heterogenität der Projekte kann dieses Wirkungsschema somit grundsätzlich für die Analyse der Evaluationsberichte herangezogen werden.

Wir bewerten die Evaluationen anhand der Güte und Vielfalt der Erfolgsindikatoren sowie der Validität der Wirkungsanalyse.⁵ Eine klassische Meta-Evaluation lässt sich aufgrund der Datenlage und -qualität leider nicht durchführen: die geringe Anzahl an Evaluationen, insbesondere belastbarer Evaluationen, sowie die hohe Heterogenität innerhalb der Fragestellungen erschweren die Vergleichbarkeit auf quantitativer Ebene.

3.1 Güte und Vielfalt der Erfolgsindikatoren

Zur Bewertung der verwendeten Erfolgsindikatoren werden die Fragen der Evaluationsstudien den drei Stufen des Logic Chart zugeordnet. Idealerweise enthalten Evaluationen Informationen zu allen Wirkungsebenen. Als unmittelbares Ergebnis einer Fördermaßnahme für Schülerinnen erfasst der Output die Akzeptanz des Projekts bei den Teilnehmerinnen. Diese lässt sich mithilfe der Bewertung der unterschiedlichen Dimensionen eines Programms durch die teilnehmenden Schülerinnen messen (Löther & Girlich, 2011). Wir ordnen alle Informationen zur Inanspruchnahme und Fragen zur Empfehlung des Programms sowie zur Bewertung der Organisation und Inhalte der Angebote dieser Wirkungsebene zu (siehe 1.1 bis 1.3 in Abbildung 1). Fragen nach der Zufriedenheit der Teilnehmerinnen, welche ein wichtiger Indikator für die Akzeptanz einer Fördermaßnahme ist, werden ebenfalls dem Output zugeordnet.

Eine Schlüsselrolle für die Aussagekraft einer Evaluation spielen die Fragen zum Outcome. Nur wenn die Teilnehmerinnen ein gesteigertes Interesse an MINT-Berufen entwickeln, nützliche Unterstützung bei der Studien- und Berufswahl erhalten, ihre Kompetenzen oder zumindest die Bewertung ihrer Kompetenzen aufbessern oder Geschlechterstereotype hinsichtlich Frauen in MINT-Berufen abbauen, kann ein MINT-Projekt die spätere Studien- und Berufswahl in gewünschter Weise beeinflussen (s. 2.1 bis 2.4 in Abbildung 1). Alle Fragen, die sich auf diese Veränderungen von Einstellungen und Bewertungen der Teilnehmerinnen beziehen, ordnen wir demnach der Outcome-Ebene zu.

Häufig enthalten die Evaluationen auch Fragen zum persönlichen Nutzen des Projekts für die Teilnehmerinnen. Sofern sich diese Fragen nicht konkret auf einzelne Outcome-Indikatoren beziehen, bleibt unklar, ob der persönliche Nutzen der Schülerinnen dazu beitragen kann, dass sie mit höherer Wahrscheinlichkeit ein Studienfach oder einen Beruf im MINT-Bereich wählen werden. Der Nutzen der Teilnahme könnte letztlich darin bestehen, dass die Schülerinnen nach Abschluss des Projekts wissen, dass sie *keinen* MINT-Beruf ergreifen wollen oder dass sie einen persönlichen Nutzen in dem neu erworbenen Wissen se-

5 Weitere Qualitätskriterien von Evaluationsstudien, wie beispielsweise Informationen zur Kompetenz der Evaluatoren, der Interaktion zwischen den Partnern des Evaluationsprozesses oder der Implementierung des Prozesses (DeGEval, 2016) können hier nicht analysiert werden, da ein Großteil der Evaluationsberichte hierzu keine Angaben machen.

hen (um beispielsweise ihre schulischen Leistungen zu verbessern). Somit werten wir allgemeine Fragen nach dem Nutzen nicht als Indikatoren für den Outcome der Maßnahme.

Analyse der Evaluationen: Da keine der vorliegenden Evaluationen die tatsächliche spätere Studien- oder Berufswahl der Schülerinnen beobachtet, kann der Impact nur anhand der Intention approximiert werden. Einige Evaluationen erfragen beispielsweise die aktuelle Neigung, einen Studiengang oder Beruf im MINT-Bereich aufzunehmen, was als Indikator für die spätere Entscheidung interpretiert werden kann, da sie die Intention zur Studien- und Berufswahl im MINT-Bereich erfassen und damit tatsächliche Verhaltensänderungen zumindest annähernd beschreiben (s. Abschnitt 2). Demnach werten wir Fragen nach den Studien- und Berufszielen als Proxy für den Impact einer Maßnahme.

Tabelle 1: Klassifizierung der Fragen

	Anzahl der Fragen pro Evaluation (Durchschnitt)	Anzahl (Anteil) der Evaluationsberichte ohne Fragen zur jeweiligen Wirkungsebene
Output	2.42	2 (15%)
Outcome	3.07	2 (15%)
Impact	1.40	5 (38%)
Insgesamt	6.89	–

Quelle: Eigene Erhebung im Rahmen des Forschungsvorhabens „MINT-Strategien 4.0“. Grundgesamtheit: 13 Evaluationsberichte.

Tabelle 1 zeigt die Zuordnung der genutzten Fragen zu den drei Wirkungsebenen. Die Evaluationen der Schülerinnen-Projekte nutzten im Mittel ungefähr sieben Fragen zur Messung des Erfolgs. Im Durchschnitt bezogen sich 2.42 Fragen pro Evaluation auf die Bewertung und Akzeptanz der konkreten Maßnahmen, und 3.07 Fragen zielten darauf ab, die verschiedenen Indikatoren auf der Outcome-Ebene (also der Determinanten der Studien- und Berufswahl) zu messen. Immerhin 1.40 Fragen pro Fragebogen waren dazu geeignet, den langfristigen Impact zu messen. Angesichts der teilweise sehr kurzen Dauer der Schülerinnen-Projekte ist es nicht verwunderlich, dass der Schwerpunkt der Fragen nicht im Bereich der Impact-Indikatoren liegt. Gleichwohl enthält eine umfassende Evaluation auch Fragen zu den tatsächlichen Verhaltensänderungen bzw. der Verhaltensintention, was bei knapp 40% der hier analysierten Evaluationen nicht der Fall ist. Das heißt, der Großteil der vorliegenden Evaluationen von Fördermaßnahmen für Schülerinnen konzentriert sich auf eine Abfrage der Einstellungsänderungen bei den Teilnehmerinnen (z.B. Steigerung des Interesses am MINT-Bereich). Intentionen zur Verhaltensänderung, also ob die Teilnehmerinnen aufgrund der Teilnahme an der Fördermaßnahme planen, eine Studien- und Berufswahl zugunsten des MINT-Bereichs zu treffen, werden jedoch am wenigsten untersucht: Fünf von 13 Berichten bzw. 38% der Evaluationsberichte erheben keine Informationen, die den Impact des jeweiligen MINT-Projekts messen, wohingegen nur 15% der Berichte keine Fragen zum Outcome stellen. Evaluationen, die keine Fragen zum Output stellen (15%), verwehren sich dadurch die Möglichkeit der inhaltlichen Nachbesserung und Anpassung ihres MINT-Projekts auf die Zielgruppe der Schülerinnen.

3.2 Validität der Wirkungsanalyse

Ein MINT-Projekt für Schülerinnen gilt dann als erfolgreich, wenn aufgrund der Teilnahme an der Maßnahme die Wahrscheinlichkeit steigt, dass die Schülerinnen einen Beruf oder ein Studium im MINT-Bereich aufnehmen und danach in diesem Beruf arbeiten. Wie in Abschnitt 2 dargestellt, kann die erhöhte Wahrscheinlichkeit anhand von veränderten Determinanten der Studien- und Berufswahl und einer stärkeren Intention, ein MINT-Studium aufzunehmen oder einen MINT-Beruf zu ergreifen, gemessen werden.

$$\delta = Y_{i1}(X_{i1}, Z_1, T = 1) - Y_{i1}(X_{i1}, Z_1, T = 0)$$

Der kausale Treatment-Effekt δ wird also als Differenz des Erfolgsindikators Y_{i1} (z.B. die Intention, ein MINT-Studium oder einen MINT-Beruf zu ergreifen, oder eine andere Outcome-Variable, gemessen zum Zeitpunkt $t = 1$) bei Teilnahme ($T = 1$) und bei Nichtteilnahme ($T = 0$) an einem Schülerinnen-Projekt gemessen. X_{i1} beschreibt weitere individuelle Einflussfaktoren zum Zeitpunkt $t = 1$ (z.B. die Schulnoten oder die geschlechtsspezifische Sozialisation) und Z_1 allgemeine Determinanten (zum Beispiel das Lohnniveau oder das Prestige eines Berufs) auf die persönliche Berufswahl. Die Herausforderung der Evaluation liegt nun darin, dass eine Person i nie in beiden Zuständen gleichzeitig beobachtet werden kann (Wooldridge, 2013). Zur Überwindung dieser Problematik lassen sich verschiedene Evaluationsdesigns nutzen, die sich unterschiedlich gut zur Kausalitätsanalyse eignen und jeweils Vor- und Nachteile in Planung, Durchführung und Aussagekraft mit sich bringen. Im Folgenden stellen wir die Möglichkeiten und Restriktionen verschiedener Untersuchungsdesigns dar und bewerten anschließend die Validität der in den vorliegenden Evaluationsberichten verwendeten Designs anhand dieser Klassifizierung. Wir beginnen mit der idealtypischen Wirkungsmessung von MINT-Schülerinnen-Projekten und diskutieren dann die eher pragmatischen Herangehensweisen.

Das experimentelle Untersuchungsdesign ist methodisch am besten zur Ursache-Wirkungs-Messung geeignet. Der kontrafaktische Zustand $Y_{i1}(X_{i1}, Z_1, T = 0)$ wird hierbei mit Hilfe einer Kontrollperson $Y_{j1}(X_{j1}, Z_1, T = 0)$ bzw. Kontrollgruppe gemessen. Diese Methode erzeugt in der Praxis jedoch nur dann valide Ergebnisse, wenn sich die Personen nicht hinsichtlich der relevanten personenspezifischen Merkmale unterscheiden ($X_{i1} = X_{j1}$). Da die Teilnehmerinnen der MINT-Projekte vermutlich aber bereits eine höhere Affinität zu diesem Fächerspektrum aufweisen als Mädchen, die nicht an einer Maßnahme teilnehmen, könnte ein einfacher Vergleich der Teilnehmerinnen mit Nicht-Teilnehmerinnen den Effekt einer Maßnahme überschätzen. Um sicherzustellen, dass sich die beiden Gruppen nicht in wesentlichen Eigenschaften voneinander unterscheiden und letztlich vergleichbar sind, müssten die Teilnehmerinnen per Zufallsauswahl auf die Gruppen verteilt werden, was im Falle der Schülerinnen-Projekte aufgrund ethischer Bedenken nicht umsetzbar ist. Theoretisch stellen mehrmalige Befragungen von Vergleichs- und Kontrollgruppe (Difference-in-difference-Ansatz) zwar eine Möglichkeit dar, die Unterschiede bei beobachtbaren und nicht-beobachtbaren Determinanten zu eliminieren und die so geschätzten Effekte präzise und unverzerrt zu messen (Wooldridge, 2013). In der Praxis ist dieses Vorgehen jedoch sehr aufwendig. Darüber hinaus können die Ergebnisse dennoch durch kleine Stichprobengrößen, den Hawthorne-Effekt oder die Unmöglichkeit von „verdeckten“ Treatments sowie Placeboeffekte (Colbjørnsen, 2003; Deaton & Cartwright, 2018) verzerrt sein.

Das quasi-experimentelle Design, das der Logik eines Experiments folgt, aber bei der Auswahl der Treatment- und Kontrollgruppe keine Randomisierung vornimmt, ist das Untersuchungsdesign mit der nächstbesten Aussagekraft zur kausalen Effektmessung. Die Auswirkungen einer fehlenden Zufallsaufteilung lassen sich zwar bei großen Stichproben mithilfe multivariater statistischer Verfahren und/oder dem Matching der Teilnehmerinnen der Versuchs- und Kontrollgruppe mildern, jedoch besteht dennoch das Risiko der Verzerrung der Effekte des Treatments durch nicht gemessene Drittvariablen (Diekmann 2012). Darüber hinaus finden MINT-Projekte typischerweise in kleinen Gruppen statt, so dass große Stichproben eher unwahrscheinlich sind. Alternativ könnte im Falle kleiner Stichproben das sequenzielle Treatment angewendet werden, bei dem die Teilnehmerinnen anderer/späterer Maßnahmen (z.B. Schülerinnen der Warteliste der Maßnahme) als Kontrollgruppe dienen (Duflo, Glennerster & Kremer, 2007). Dies setzt jedoch voraus, dass es eine Warteliste gibt oder die Teilnehmerinnen der Schülerinnen-Projekte relativ langfristig vorher bekannt sind.

Vorexperimentelle Designs verzichten auf den Vergleich mit einer Kontrollgruppe und befragen die Teilnehmerinnen entweder vor und nach der Teilnahme am Schülerinnen-Projekt oder nur nach dem Projekt (Diekmann, 2012). Da bei einer einmaligen Befragung jedoch keine Veränderungen der Determinanten der Studien- und Berufswahl gemessen werden können, lassen sich auf diesem Weg keine kausalen Effekte nachweisen⁶. Ein Vorher-Nachher-Vergleich ohne Kontrollgruppe kann dagegen nur dann den kausalen Effekt einer Maßnahme auf die Determinanten der Berufswahl messen, wenn der Outcome nach Nichtteilnahme $Y_{i1}(X_{i1}, Z_1, T = 0)$ identisch ist mit dem Outcome vor Nichtteilnahme $Y_{i0}(X_{i0}, Z_0, T = 0)$. Theoretisch kann aber nicht ausgeschlossen werden, dass sich während des Beobachtungszeitraums die Studien- und Berufswahlentscheidungen insgesamt gewandelt haben und die Veränderungen, die bei den Teilnehmerinnen beobachtet werden, nicht durch das Projekt, sondern durch Veränderungen der persönlichen Einflussfaktoren X_i oder allgemeine Trends (Veränderungen der Determinanten Z) ausgelöst wurden.⁷ Je kürzer der Beobachtungszeitraum und stabiler die sonstigen Determinanten Z der Berufswahl sind, desto geringer ist jedoch diese Problematik.

Da die Schülerinnen-Projekte zumeist auf einen oder wenige Tage begrenzt sind, erscheint der Nutzen eines Kontrollgruppendedesigns (mit ein- oder mehrmaliger Befragung) bei der Evaluation von MINT-Schülerinnen-Projekten eher fraglich. Darüber hinaus sind die persönlichen Einflussfaktoren X_i (z.B. die persönlichen Kompetenzen, die geschlechtsspezifische Sozialisation der Schülerinnen und dazugehörige Einflüsse wie die Einstellungen von Freunden und Familie) zumeist zeitlich sehr stabil, so dass während des relativ kurzen Beobachtungszeitraums keine gravierenden Veränderungen zu erwarten sind und ein einfacher Vorher-Nachher-Vergleich durchaus zu vergleichsweise verlässlichen Ergeb-

6 Folgt man den Befunden von Müller (2016), zeigt sich jedoch, dass eine retrospektive Selbsteinschätzung nur unter bestimmten Bedingungen eine valide Messung der kontrafaktischen Situation darstellt.

7 Dies wäre beispielweise der Fall, wenn sich während des Beobachtungszeitraums die Löhne in den MINT-Berufen drastisch erhöhen, die Vereinbarkeit von Familie und Beruf offensichtlich verbessern und die Arbeitslosenquote in den frauendominierten Berufen signifikant erhöhen würde, so dass sich Frauen aufgrund dieser Veränderungen ohnehin deutlich häufiger für MINT-Berufe entscheiden würden.

nissen führen kann. Dieses Untersuchungsdesign lässt sich auch in der Praxis ohne größeren Aufwand umsetzen und stellt daher eine gute und pragmatische Lösung dar.

Nichtsdestotrotz müssen die Befragungsergebnisse mit Vorsicht interpretiert werden. So erhöht die häufig direkt nach der Teilnahme an der Maßnahme stattfindende Befragung der Schülerinnen die Wahrscheinlichkeit sozial erwünschter Antworten. Dies gilt insbesondere dann, wenn die Teilnehmerinnen die Fragebögen in Anwesenheit der Projektkoordinatorinnen und Projektkoordinatoren ausfüllen, deren Projekt sie möglicherweise nicht schlecht bewerten möchten. Auch wenn die Schülerinnen dies nicht intendieren, führt der Hawthorne-Effekt, der durch die besondere Aufmerksamkeit entsteht, die den Teilnehmerinnen zuteilwird, bei einer Prä-Post-Befragung zu einer Überschätzung der gemessenen Effekte (Colbjørnsen, 2003). Diese Verzerrung obliegt jedoch allen Befragungen und kann nur durch die Gestaltung einer professionellen Befragungssituation gemildert werden. Zum anderen wäre es denkbar, dass die Wirkung der Fördermaßnahme direkt nach der letzten Veranstaltung größer ist als zu einem späteren Zeitpunkt, so dass insbesondere die langfristigen Effekte eventuell überschätzt werden.

Aus unserer Sicht liefern somit Vorher-/Nachher-Befragungen auch ohne den Vergleich mit einer Kontrollgruppe durchaus valide Informationen, um Veränderungen in den Einstellungen und Intentionen der Teilnehmerinnen von MINT-Projekten erfassen zu können.

Analyse der Evaluationen: In den vorliegenden Evaluationen werden nur zwei verschiedene Untersuchungsdesigns angewendet. Unsere erste Kategorie umfasst alle Einmalbefragungen am Ende des Projekts. In Kategorie 2 fallen Evaluationen, deren Befragte sowohl vor als auch nach dem Projekt Fragebögen beantworteten. Evaluationen aus der Kategorie 1 erlauben aus wissenschaftlicher Sicht keine Schlussfolgerungen über die kausale Wirkung der Maßnahme.

Von den 13 vorliegenden Evaluationsberichten führte der Großteil der Fördermaßnahmen einmalige Befragungen zum Ende des MINT-Projekts durch (Kategorie 1). Nur drei der analysierten Evaluationen beschreiben, dass Vorher-Nachher-Befragungen (Kategorie 2) genutzt wurden, und werden damit zumindest den minimalen Anforderungen an eine Evaluation eines MINT-Schülerinnen-Projekts gerecht. Die Validität der Wirkungsanalyse der existierenden Evaluationen ist in der Summe demnach eher begrenzt. Angesichts der deutlich komplexeren Datenerhebung und der begrenzten Vorteile hinsichtlich der Validität ist es nicht verwunderlich, dass es unseres Wissens keine Evaluation mit Hilfe eines Kontrollgruppendesigns gibt.

4 Zusammenfassung und kritische Reflexion

Hochschulen für angewandte Wissenschaften versuchen durch gezielte Angebote für Schülerinnen den Frauenanteil in MINT-Studiengängen zu erhöhen. Angesichts der bundesweit nur sehr geringfügigen Erhöhung der Frauenanteile und der beträchtlichen personellen und finanziellen Ressourcen für die Durchführung dieser Aktivitäten stellt sich dringend die Frage nach der Wirksamkeit dieser Maßnahmen.

Auch wenn kein direkter kausaler Zusammenhang zwischen der Teilnahme an einer Maßnahme und der späteren Berufswahl aufgrund der Komplexität der Entscheidung zu

erwarten ist, so sollten die Schülerinnen-Projekte zumindest dazu beitragen, die Hürden zur Aufnahme eines MINT-Studiums oder -Berufs abzubauen und/oder die Determinanten dieser Bildungsentscheidung positiv beeinflussen. Ob die Schülerinnen-Projekte tatsächlich einen Einfluss auf die relevanten und veränderbaren Einflussfaktoren der Studien- und Berufswahl haben, lässt sich allerdings anhand der uns vorliegenden Evaluationsberichte nicht beantworten. Unsere Analysen zeigen, dass die bisherigen Evaluationen sowohl hinsichtlich der Güte und Vielfalt der Erfolgskriterien als auch der Validität der Wirkungsanalyse verbesserungsfähig sind. Die Bewertung eines Projekts anhand einer Selbstauskunft der teilnehmenden Schülerinnen zur Zufriedenheit mit dem Projekt (Output) lässt keine Schlüsse über die Effekte auf die Determinanten der Studien- und Berufswahl zu. Hierzu müssten grundsätzlich mehr Fragen zu den verschiedenen Outcome-Indikatoren gestellt werden. Auch von der Möglichkeit, den langfristigen Impact anhand der Intention der Studien- und Berufswahl zu approximieren, wird sehr selten Gebrauch gemacht. Darüber hinaus lassen sich kausale Effekte auf Einstellungen der Schülerinnen nicht mithilfe der meistens verwendeten Einmalbefragung am Ende des Projekts feststellen.

Auch wenn die Datenbasis unserer Meta-Evaluation sehr gering ist, gehen wir davon aus, dass die Ergebnisse ein realistisches Bild der vorhandenen Evaluationen abgeben. Aus der Kommunikation mit Projektverantwortlichen hat sich gezeigt, dass einige Projektverantwortliche und/oder Evaluierende, z.B. aus Sorge um Datenschutz und Wahrung der Anonymität der Teilnehmerinnen, ihre Evaluationsberichte nicht zur Verfügung stellen wollten. Insbesondere im Hinblick darauf, dass die Projekte häufig von derselben Person durchgeführt und evaluiert werden, steigt der Anreiz für die Verantwortlichen, unvoreilhaftige Ergebnisse nicht zu veröffentlichen. Da wissenschaftlich profundere Studien vermutlich eher von externen Wissenschaftlerinnen und Wissenschaftlern durchgeführt werden, ist nicht unbedingt zu erwarten, dass uns diese Evaluationen systematisch vorenthalten wurden.

Die relativ kleine Anzahl an auswertbaren Evaluationen schmälert aus unserer Sicht auch nicht die Erkenntnisse aus dieser Meta-Evaluation. Ganz im Gegenteil, sie verdeutlicht die Relevanz unserer Forschung und macht das große Potenzial im Bereich der Evaluationen von MINT-Projekten für Schülerinnen an HAW deutlich. Es mangelt nicht nur an der Anzahl empirischer Studien, sondern auch an der wissenschaftlichen Fundierung der Evaluationsansätze. Angesichts dieser Wissenslücke stehen viele Hochschulen vor der offenen Frage, wie sie ihre knappen Ressourcen für die Anwerbung von Frauen in MINT-Studiengängen am effektivsten nutzen können. Darüber hinaus könnten aussagekräftige Evaluationen wichtigen Input für die Weiterentwicklung von Maßnahmen liefern.

Das von uns entwickelte Wirkungsmodell, welches explizit die drei Wirkungsebenen von MINT-Angeboten für Schülerinnen an HAW systematisch darstellt, sowie die Diskussion der verschiedenen Wirkungsanalysen können Projektverantwortlichen auch bei der Konzeption einer eigenen Evaluation dienlich sein. Darüber hinaus ist ein „Werkzeugkasten“ mit praxistauglichen Anleitungen zur Durchführung von Evaluationen hilfreich, um methodisches Wissen aus der Wissenschaft an die Praxis weiterzugeben und somit die Erstellung belastbarer Evaluationen von laufenden und zukünftigen MINT-Projekten zu fördern (siehe z.B. Nickolaus & Mokhonko, 2016; Nationales MINT Forum, 2018; sowie Brenning & Wolf, 2021).

Literatur

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t)
- Anger, C., Koppel, O. & Plünnecke, A. (2018). *MINT-Frühjahrsreport 2018. MINT – Offenheit, Chancen, Innovationen*. Verfügbar unter: [https://www.arbeitgeber.de/www/arbeitgeber.nsf/res/MINT-Fruehjahrensreport_2018.pdf/\\$file/MINT-Fruehjahrensreport_2018.pdf](https://www.arbeitgeber.de/www/arbeitgeber.nsf/res/MINT-Fruehjahrensreport_2018.pdf/$file/MINT-Fruehjahrensreport_2018.pdf)
- Armitage, C. J. & Conner, M. (2001). Efficacy of the theory of planned behaviour. A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471–499. <https://doi.org/10.1348/014466601164939>
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: Freeman. <https://doi.org/10.4135/9780857025753.n178>
- Bandura, A., Barbaranelli, C., Caprara, G. V. & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. <https://doi.org/10.1111/1467-8624.00273>
- BLK – Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (2002). *Frauen in den ingenieur- und naturwissenschaftlichen Studiengängen. Bericht der BLK vom 2. Mai 2002. Materialien zur Bildungsplanung und zur Forschungsförderung, Heft 100*. Bonn.
- Brenning, S. & Wolf, E. (2019). *What works? Eine Meta-Evaluation von Evaluationen von MINT-Projekten für Schülerinnen und Studentinnen*. Verfügbar unter: https://www.oth-regensburg.de/fileadmin/media/fakultaeten/s/forschung_projekte/mint/MINT-Strategien_4_0_Meta-Evaluation_EW_SB.pdf
- Brenning, S. & Wolf, E. (2021). *Wirkung messen. Handbuch zur Evaluation von MINT-Projekten für Schülerinnen*. Hochschule München.
- Colbjørnsen, T. (2003). Der Hawthorne-Effekt oder die Human-Relations-Theorie. Über die experimentelle Situation und ihren Einfluss. In S. U. Larsen & E. Zimmermann (Hrsg.), *Theorien und Methoden in den Sozialwissenschaften* (S. 131–143). Wiesbaden: Westdeutscher Verlag. https://doi.org/10.1007/978-3-322-80451-8_10
- Correll, S. J. (2001). Gender and the career choice process. The role of biased self-assessments. *American Journal of Sociology*, 106(6), 1691–1730. <https://doi.org/10.1086/321299>
- Correll, S. J. (2004). Constraints into preferences. Gender, status, and emerging career aspirations. *American Sociological Review*, 69(1), 93–113. <https://doi.org/10.1177/000312240406900106>
- Deaton, A. & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210, 2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- DeGEval – Gesellschaft für Evaluation e.V. (2016). *Standards für Evaluation*. Erste Revision 2016. Mainz: DeGEval – Gesellschaft für Evaluation e.V.
- Diekmann, A. (2012). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Heidelberg: Springer.
- Duflo, E., Glennerster, R. & Kremer, M. (2007). *Using randomization in development economics research. A toolkit*. Discussion Paper Nr. 6059. Centre for Economic Policy Research, London.
- Ertl, B., Luttenberger, S. & Paechter, M. (2014). Stereotype als Einflussfaktoren auf die Motivation und die Einschätzung der eigenen Fähigkeiten bei Studentinnen in MINT-Fächern. *Gruppendynamik und Organisationsberatung*, 45(4), 419–440. <https://doi.org/10.1007/s11612-014-0261-3>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Hiller, S. (2014). *Ergebnisbericht zur wissenschaftlichen Evaluation von EnterTechnik*. Stuttgart: Dialogik gGmbH.

- Hurrelmann, K., Bauer, U., Grundmann, M. & Walper, S. (2015). *Handbuch Sozialisationsforschung*. Weinheim: Beltz.
- Jones, M. G., Howe, A. & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education*, 84(2), 180–192. [https://doi.org/10.1002/\(sici\)1098-237x\(200003\)84:2<180::aid-sce3>3.0.co;2-x](https://doi.org/10.1002/(sici)1098-237x(200003)84:2<180::aid-sce3>3.0.co;2-x)
- Kenkmann, T. (2005). 5. *Brandenburgische Sommer-Universität für Schülerinnen in Naturwissenschaft und Technik*. Abschlussbericht.
- Kessels, U. & Hannover, B. (2006). Zum Einfluss des Image von mathematisch-naturwissenschaftlichen Schulfächern auf die schulische Interessenentwicklung. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule*. Abschlussbericht des DFG-Schwerpunktprogramms (S. 350–369). Münster: Waxmann.
- Kessels, U. (2015). Bridging the gap by enhancing the fit: How stereotypes about STEM clash with stereotypes about girls. *International Journal of Gender, Science and Technology*, 7(2), 280–296.
- Kim, M. & Hunter, J. E. (1993). Relationships among attitudes, behavioral intentions, and behavior. A meta-analysis of past research, part 2. *Communication Research*, 20, 331–364. <https://doi.org/10.1177/009365093020003001>
- Kray, J. (2019). *Entwicklungspsychologie. Ein Überblick für Psychologiestudierende und -interessierte*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-57761-5>
- Lent, R. W., Brown, S. D. & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior*, 45, 79–122. <https://doi.org/10.1006/jvbe.1994.1027>
- Löther, A. & Girlich, J. (2011). Frauen in MINT-Fächern: Bilanzierung der Aktivitäten im hochschulischen Bereich. *Materialien der Gemeinsamen Wissenschaftskonferenz. Heft 21*. Bonn.
- Matthes, S. (2019). *Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions- und Aversionsfaktoren in der Berufsfindung*, Institut für Berufsbildung, Bonn.
- Mokhonko, S., Nickolaus, R. & Windaus, A. (2014). Förderung von Mädchen in Naturwissenschaften. Schülerlabore und ihre Effekte. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 143–159. <https://doi.org/10.1007/s40573-014-0016-2>
- Mokhonko, S. (2016). *Nachwuchsförderung im MINT-Bereich. Aktuelle Entwicklungen, Fördermaßnahmen und ihre Effekte*. Stuttgart: Franz Steiner.
- Müller, C. E. (2016). Wirkungsbewertung mittels retrospektiver Pretests: Eine kritische Würdigung. *Zeitschrift für Evaluation*, 15(2), 221–239.
- Müller, C. & Albrecht, M. (2016). The future of impact evaluation is rigorous and theory-driven. In R. Stockmann & W. Meyer (Hrsg.), *The future of evaluation. Global trends, new challenges, shared perspectives* (S. 283–293). New York: Palgrave Macmillan. https://doi.org/10.1057/9781137376374_21
- Nationales MINT Forum (2018). *Orientierungsrahmen für wirkungsvolle Arbeit außerschulischer MINT-Initiativen*. Ein praktischer Leitfaden zur Selbstanalyse. Berlin. Verfügbar unter: https://www.nationalesmintforum.de/fileadmin/medienablage/content/publikationen_und_empfehlungen/publikationen/2018/NMF_Leitfaden_Web.pdf
- Nickolaus, R. & Mokhonko, S. (Hrsg.) (2016). *In fünf Schritten zum zielführenden Evaluationsdesign. Eine Handreichung für Bildungsinitiativen im MINT-Bereich* (acatech Materialien). München: acatech – Deutsche Akademie der Technikwissenschaften. Verfügbar unter: <https://www.acatech.de/publikation/in-fuenf-schritten-zum-zielfuehrenden-evaluationsdesign-eine-handreichung-fuer-bildungsinitiativen-im-mint-bereich/>
- Obergruber, N. (2018). *Microeconomic analysis of individual and institutional determinants of education and occupational choice*. Inaugural-Dissertation zur Erlangung des Grades Doctor oeco-

- nomiae publicae (Dr. oec. publ.) an der Ludwig-Maximilians-Universität München.
<https://doi.org/10.5282/edoc.22892>
- Patton, M. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks: Sage.
- Pöllmann-Heller, K. & Rudolph, C. (2019). Frauen in das MINT-Studium. Ambivalenzen und Potenziale von hochschulischen Förderprozessen. In C. Driesen & A. Ittel (Hrsg.), *Der Übergang in die Hochschule. Strategien, Organisationsstrukturen und Best Practices an deutschen Hochschulen* (S. 91–102). Münster: Waxmann.
- Prenzel, M., Lankes, E. & Minsel, B. (2000). Interessenentwicklung in Kindergarten und Grundschule. Die ersten Jahre. In U. Schiefele & K.-P. Wild (Hrsg.), *Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Motivation* (S. 11–30). Münster: Waxmann.
- Randall, D. M. & Wolff, J. A. (1994). The time interval in the intention-behaviour relationship. Meta-analysis. *British Journal of Social Psychology*, 33, 405–418.
<https://doi.org/10.1111/j.2044-8309.1994.tb01037.x>
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, 2(5), 36–38.
- Shapiro, J. R. & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66, 175–183.
<https://doi.org/10.1007/s11199-011-0051-0>
- Sheu, H., Lent, R. W., Brown, S. D., Miller, M. J., Hennessy, K. D. & Duffy, R. D. (2010). Testing the choice model of social cognitive career theory across Holland themes: A meta-analytic path analysis. *Journal of Vocational Behavior*, 76(2), 252–264.
<https://doi.org/10.1016/j.jvb.2009.10.015>
- Siegler, R., Eisenberg, N., De Loache, J. & Saffran, J. (2016). *Entwicklungspsychologie im Kindes- und Jugendalter*. Heidelberg: Springer. https://doi.org/10.1007/978-3-662-47028-2_16
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
<https://doi.org/10.1006/jesp.1998.1373>
- Solga, H. & Pfahl, L. (2009). Doing Gender im technisch-naturwissenschaftlichen Bereich. In J. Milberg (Hrsg.), *Förderung des Nachwuchses in Technik und Naturwissenschaft* (S. 155–218). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-01123-8_4
- Statistisches Bundesamt (2019). *DESTATIS – Bildung, Forschung, Kultur: Studierende in MINT-Fächern*. Verfügbar unter:
<https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-mint-faechern.html>
- Stöger, H., Duan, X., Schirmer, S., Greindl, T. & Ziegler, A. (2013). The effectiveness of a one-year online mentoring program for girls in STEM. *Computers & Education*, 69, 408–418.
<https://doi.org/10.1016/j.compedu.2013.07.032>
- Taskinen, P. (2010). *Naturwissenschaften als zukünftiges Berufsfeld für Schülerinnen und Schüler mit hoher naturwissenschaftlicher und mathematischer Kompetenz: Eine Untersuchung von Bedingungen für Berufserwartungen* (Dissertation). Christian-Albrechts-Universität zu Kiel, Kiel.
- Taskinen, P. H., Schütte, K. & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career. The role of school factors, individual interest and science self-concept. *Educational Research and Evaluation*, 19(8), 717–733.
<https://doi.org/10.1080/13803611.2013.853620>
- Weinhardt, F. (2017). Ursache für Frauenmangel in MINT-Berufen? Mädchen unterschätzen schon in der fünften Klasse ihre Fähigkeiten in Mathematik. *DIW Wochenbericht*, 45/2017.
- Wentzel, W. (2008). „Ich will das und das ist mein Weg!“ *Junge Frauen auf dem Weg zum Technikberuf. Qualitative Interviews mit ehemaligen Girls' Day-Teilnehmerinnen in Ausbildung und Studium*. Bielefeld: Kompetenzzentrum Technik – Diversity – Chancengleichheit.

- Widmer, T. (1996). *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.
- W. K. Kellogg Foundation (2004). *Logic Model Development Guide. Using Logic Models to bring together Planning, Evaluation and Action*. Michigan. Verfügbar unter:
<https://www.bttop.org/sites/default/files/public/W.K.%20Kellogg%20LogicModel.pdf>
- Wooldridge, J. (2013). *Introductory econometrics: A modern approach*. Mason: Nelson Education.
- Zeisberg, I. (2013). Light up your life. Einfluss geschlechtsspezifischer Maßnahmen in außerschulischen Lernorten auf MINT-Interesse und Berufswahl. *Wissenschaftliche Schriften der WWU Münster*, Reihe VI, 12.

Kontakt:

Stefanie Brenning, M.A. · Prof. Dr. Elke Wolf
Hochschule München
Lothstraße 64
80335 München
E-Mail: s_brenning@gmx.de
E-Mail: elke.wolf@hm.edu