

---

**Hinweis:** Die Nutzer dürfen die Inhalte nur zum Zwecke der wissenschaftlichen Forschung ansehen, drucken, kopieren, herunterladen, sowie für Text- und Datamining verwenden. Die Inhalte dürfen weder ganz noch teilweise wörtlich (wieder)veröffentlicht oder für kommerzielle Zwecke verwendet werden. Die Nutzer müssen sicherstellen, dass die Urheberpersönlichkeitsrechte des Autors sowie gegebenenfalls bestehende Rechte Dritter an den Inhalten oder Teilen der Inhalte nicht verletzt werden.

Dies ist das akzeptierte Manuskript (Author's Accepted Manuscript) des folgenden Beitrages:  
Weis, S., Stoll, G., Dietrich, S. & Wachtel, T., Evaluation von Online Self-Assessments, veröffentlicht in *Online Self-Assessments zur Studienfachwahl: Entwicklung - Konzepte - Qualitätsstandards*, herausgegeben von Stoll, G. & Weis, S., 2021, Springer Verlag, vervielfältigt mit Genehmigung von: Springer Verlag.

## 9 Evaluation von Online Self-Assessments

*Susanne Weis, Gundula Stoll, Sylvia Dietrich & Tatjana Wachtel*

### Leitfragen

1. Welche methodischen Entscheidungen und Schritte sind im Evaluationsprozess relevant?
2. Welche Kriterien sind für die Evaluation der Wirksamkeit von OSA relevant und geeignet?
3. In welchem Verhältnis stehen die Kriterien untereinander?
4. Welche versuchsplanerischen Überlegungen sind bei der Wahl eines adäquaten Forschungsdesigns zur Untersuchung der Wirksamkeit von OSA maßgeblich?

### 9.1 Überblick

Mit der Evaluation von OSA wird häufig die Abschlussbefragung von Nutzerinnen und Nutzern nach erfolgter Teilnahme verstanden. Dies erfolgt meist mit dem Ziel, die Zufriedenheit und Akzeptanz des OSA nach außen zu dokumentieren und ggf. Stellschrauben zu identifizieren, um das OSA im Hinblick auf seine Nutzerfreundlichkeit zu verbessern. Auch empirische Studien zur Überprüfung der Vorhersagekraft der eingesetzten diagnostischen Elemente für den späteren Studienerfolg (s. Kap. 4 und 5) können als Evaluation von OSA betrachtet werden. Dass eine Evaluation von OSA allerdings wesentlich umfassender sein kann und sollte, und welche Aspekte dabei berücksichtigt werden müssen, soll das vorliegende Kapitel aufzeigen.

Dieses Kapitel betrachtet aus evaluationstheoretischer Sicht OSA als eine komplexe, auf verschiedenen Wirkmechanismen (s. Kap. 1) beruhende Intervention in den Prozess der Studienwahl und ist in drei Schwerpunkte gegliedert. Wir beginnen mit einem methodischen Überblick über den Evaluationsprozess (s. Abschn. 9.2) von der Klärung der Rahmenbedingungen bis zur Frage der Berichtstellung. Damit soll ein Rahmen bereitgestellt werden, anhand dessen Evaluationsstudien geplant und durchgeführt werden können. Im zweiten Schwerpunkt (s. Abschn. 9.3) befassen wir uns mit der Frage, woran man die Wirksamkeit der Intervention OSA festmachen kann, d. h. welche Kriterien für die Evaluation von OSA geeignet und sinnvoll sind. Im dritten Schwerpunkt (s. Abschn. 9.4) beleuchten wir darauf aufbauend die Frage, was bei der Untersuchungsplanung zur Überprüfung der Wirksamkeit von OSA beachtet werden muss.

Die Evaluation ist nicht an eine bestimmte Entwicklungsphase von OSA gebunden, sondern wird im Idealfall von Beginn an mitgedacht und kontinuierlich verfolgt. Das vorliegende Kapitel steht daher in Bezug zu allen übrigen Kapiteln dieses Buches und richtet sich explizit auch an Evaluatorinnen und Evaluatoren von OSA, die nicht selbst an der Entwicklung von OSA beteiligt sind. Gerade bei Vorhaben, die durch öffentliche Gelder finanziert werden – wie es bei OSA-Projekten an Hochschulen häufig der Fall ist – ist es inzwischen Standard, die Evaluation des Vorhabens direkt mit der Antragstellung zu berücksichtigen.

---

## Allgemeines zum Begriff Evaluation

### Definition

Evaluation ist eine „wissenschaftliche Dienstleistung, die namentlich öffentlich verantwortete und/oder finanzierte ‚Gegenstände‘ (Politiken, Programme, Projekte, Maßnahmen ...) in verschiedensten Themenfeldern systematisch, transparent und datengestützt beschreibt und ausgewogen bewertet, so dass Beteiligte und Betroffene (Auftraggebende und andere Stakeholder [im Folgenden Interessensträger, Anm. der Autorinnen]) die erzeugten Evaluationsergebnisse für vorgesehene Zwecke wie Rechenschaftslegung, Entscheidungsfindung oder Optimierung nutzen.“ (Beyvl & Widmer, 2009, S. 16)

Evaluation ist nach DeGEval – Gesellschaft für Evaluation e.V. (2017) gekennzeichnet durch:

1. Systematik, Transparenz, datengestütztes Vorgehen und kriteriengeleitete Bewertung: Die datengestützte Bewertung eines Evaluationsgegenstands soll anhand systematischer Verfahren und offen gelegter Kriterien erfolgen, um eine nachvollziehbare, vom Individuum unabhängige Prüfung zu ermöglichen (s. auch Stockmann, 2004).
2. Zweckgebundenheit: Evaluationen sollen für die Interessensträgerinnen und -träger und für den jeweiligen Evaluationszweck nutzbare Ergebnisse liefern.
3. Interdisziplinarität und Dienstleistung: Evaluation ist auf die unterschiedlichsten Evaluationsgegenstände (z. B. politische Entscheidungen, Produkte, Prozesse oder Maßnahmen/Programme) in den verschiedensten Anwendungsbereichen (z. B. Pflege, Medizintechnik, Forstwirtschaft etc.) anwendbar und wird auch von Personen aus ebenso unterschiedlichen Fachrichtungen durchgeführt. Damit im Einklang sollten Prozess und Ergebnisse so kommuniziert werden, dass sie für Fachfremde verständlich sind.

Eine umfangreiche Orientierungshilfe vor allem für Evaluatorinnen und Evaluatoren bieten die Standards für Evaluation, die das „übergreifende Ziel der Qualitätssicherung und -entwicklung von Evaluationen“ verfolgen (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 12). Die Standards sind in vier Kategorien unterteilt, welche vier grundlegende Eigenschaften von Evaluationen bezeichnen:

- Nützlichkeit: „Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an abgestimmten und geklärten Evaluationszwecken sowie soweit möglich am Informationsbedarf der vorgesehenen Nutzerinnen und Nutzer ausrichtet.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 17)
- Durchführbarkeit: „Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 22)
- Fairness: „Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit allen beteiligten und betroffenen Personen und Gruppen umgegangen wird.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 24)
- Genauigkeit: „Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige und nachvollziehbare Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt. In diesem Zusammenhang kommt wissenschaftlichen Gütekriterien eine besondere Bedeutung zu.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 27)

Die Evaluationsstandards gelten als Leitlinien ohne normativen Charakter. Sie bieten eine Orientierungshilfe bei der Initiierung, Planung und Durchführung von Evaluationen, können aber auch für eine Meta-Evaluation zur Beurteilung der Qualität einer Evaluation genutzt werden. Eine Ausnahme stellen die Fairness-Standards dar, die v. a. ethische und rechtliche Aspekte beinhalten und denen unter allen Umständen entsprochen werden sollte. Eine vollständige Darstellung der Standards in der jeweils gültigen Fassung ist auf den Webseiten der Deutschen Gesellschaft für Evaluation frei zugänglich (DeGEval – Gesellschaft für Evaluation e.V., 2017, <https://www.degeval.org/degeval-standards/standards-fuer-evaluation/>).

---

## 9.2 Der Evaluationsprozess

### 9.2.1 Klärung der Rahmenbedingungen

#### Welchen Zweck verfolgt die Evaluation?

„Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, [...].“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 18, N 2 Klärung der Evaluationszwecke). Mögliche Evaluationszwecke sind unter anderem die Optimierung oder Verbesserung eines Programms und die Rechenschaftslegung (s. Definition von Beywl & Widmer, 2009). Häufig sind Evaluationszwecke an eine bestimmte Entwicklungsphase und an bestimmte so genannte „Evaluationsansätze“ geknüpft:

- a. Der Zweck der Optimierung und Verbesserung eines OSA wird schwerpunktmäßig während der Entwicklung eines OSA verfolgt und ist geknüpft an den Ansatz der „formativen“ Evaluation. Die grundlegende Frage, mit der sich die Evaluation in diesem Fall beschäftigt ist die nach möglichen Verbesserungen eines OSA (s. Abschn. 9.2.2). Auch nach Abschluss einer OSA-Entwicklung kann eine „prozessbegleitende“ Evaluation durch Befragung von Teilnehmenden den Zweck der OSA-Optimierung verfolgen.
- b. Die Rechenschaftslegung erfolgt zum Abschluss eines OSA-Entwicklungsvorhabens. Sie soll Aufschluss darüber geben, ob durch das OSA verfolgte Ziele auch tatsächlich erreicht worden sind (z. B. zur Frage, ob sich der Studienerfolg der OSA-Absolventinnen und -Absolventen in der Studieneingangsphase durch das OSA steigert (s. Abschn. 9.2.2).

#### In welchem Bezug steht die Evaluation zum OSA (intern vs. extern)?

Eine externe Evaluation erfolgt durch eine Person oder Gruppe, die in keiner Beziehung zum Evaluationsgegenstand steht. Eine interne Evaluation erfolgt durch eine Person oder Gruppe, die direkt am Evaluationsgegenstand beteiligt ist (z. B. das Entwicklungsteam des OSA) (Döring & Bortz, 2016). Interne und externe Evaluation sind mit jeweils entgegengesetzten Vor- und Nachteilen verbunden. Zum Zweck der Optimierung eines OSA liegt der Vorteil der internen Evaluation in der auf das OSA bezogenen Expertise, der Nähe zu den OSA-bezogenen Tätigkeiten und der damit verbundenen unmittelbaren Möglichkeiten zur Umsetzung von Erkenntnissen und Änderungen durch entsprechend kurze Kommunikationswege innerhalb der Arbeitsgruppe. Zum Zweck der Rechenschaftslegung (z. B. bei der Evaluation der Wirksamkeit eines OSA) liegt der Vorteil der externen Evaluation in der Trennung von evaluations- und OSA-bezogenen Tätigkeiten, der Expertise in Evaluationsmethoden und damit auch in der wahrgenommenen Unabhängigkeit der Bewertung, die zu einer höheren Glaubwürdigkeit der Evaluationsergebnisse beiträgt.

Wenn ein OSA ausschließlich intern evaluiert wird, sollten folgende Maßnahmen ergriffen werden, um Zweifel an der Glaubwürdigkeit von vorneherein auszuräumen:

- Dokumentation aller evaluationsbezogenen Entscheidungen und Tätigkeiten von Beginn an und dabei Entkopplung von OSA- und evaluationsbezogenen Überlegungen
- Anwendung von prospektiven und ‚starken‘, d. h. aussagekräftigen Versuchsplänen zur Überprüfung der Wirksamkeit (s. Abschn. 9.4)
- Inanspruchnahme von (externer) methodischer Beratung
- Transparenz im Umgang mit Limitationen im Rahmen der Meta-Evaluation (s. Abschn. 9.2.5)

## Wer sind die am OSA und dessen Evaluation beteiligten Interessensträgerinnen und -träger und was sind ihre Rollen?

Idealerweise sollten die unterschiedlichen Interessengruppen bereits bei der Zielsetzung und Konzeption des OSA identifiziert und einbezogen werden (s. Kap. 1 und 2). Wenn dies nicht der Fall ist, so stellt die Identifikation der Beteiligten und Betroffenen einen ersten Schritt in der Planung des Evaluationsvorhabens dar. Die Kenntnis aller beteiligten Interessensgruppen und Betroffenen ist auch eine zentrale Voraussetzung für die spätere Nutzung der Evaluationsergebnisse durch ebendiese Personen (DeGEval – Gesellschaft für Evaluation e.V., 2017, N 1 Identifizierung der Beteiligten und Betroffenen, N 8 Nutzung und Nutzen der Evaluation; Döring & Bortz, 2016). Tab. 9.1 nennt die wesentlichen Rollen im Rahmen eines Evaluationsvorhabens und zeigt je nach Evaluationszweck auf, welche Personengruppen typischerweise welche Rollen einnehmen.

Tab. 9.1 Rollen bei der Evaluation eines OSA

	Beteiligte an der Evaluation zum Zweck der...	
Rolle	Rechenschaftslegung	Optimierung
Auftraggeber, Geldgeber	Politische Entscheidungsträger, Hochschulleitungen, Studiengänge*	Politische Entscheidungsträger, Hochschulleitungen, Studiengänge* (bei externer Evaluation auch OSA-Entwicklerinnen und -Entwickler)
Betroffene	Studieninteressierte, Studiengänge*, Lehrende an Hochschulen, Studienberatungen und Studienfachberatungen	
Datenbeschaffende	SuS, Studierende, Lehrende an Hochschulen, Studiengänge*	SuS, Studierende, Lehrende an Hochschulen, Studiengänge*
Nutzerinnen und Nutzer d. Evaluationsergebnisse	Politische Entscheidungsträger, Berufsberatungsinstitutionen, Hochschulleitungen, Studiengänge*, Lehrende an Schulen**, SuS	OSA-Entwicklerinnen und -Entwickler
Durchführende	Evaluationsteam	OSA-Entwicklerinnen und -Entwickler

\* je nach Bedeutung sind hier auch Studiengangverantwortliche gemeint

\*\* in ihrer Funktion als Multiplikatoren beim Hinweis auf geeignete OSA-Plattformen

SuS: Schülerinnen und Schüler

## Wie ist der zeitliche Rahmen? Wie und an wen wird berichtet?

Um eine realistische Planung und Umsetzung der Evaluation zu gewährleisten, müssen die zeitlichen Rahmenparameter zu Beginn geklärt werden (DeGEval – Gesellschaft für Evaluation e.V., 2017, N 7 Rechtzeitigkeit der Evaluation). Dazu zählen vor allem die Fragen, wann in Bezug auf die Entwicklung des OSA die Evaluation geplant wird und wann welche Ergebnisse in welcher Form vorliegen müssen. Nur bei ausreichendem zeitlichem Vorlauf ist es beispielsweise möglich, notwendige Messinstrumente zu entwickeln und ein passendes Versuchsdesign und geeignete Stichproben auszuwählen (s. Abschn. 9.4).

## Was ist der Evaluationsgegenstand?

Ein Evaluationsgegenstand ist das Objekt, welches bewertet werden soll (Döring & Bortz, 2016). Der Evaluationsgegenstand kann entweder das OSA als Ganzes sein, oder die verschiedenen Elemente eines OSA (z. B. diagnostische Elemente). So könnte man einerseits global den Nutzen, die Wirkung oder die Effektivität des gesamten OSA untersuchen oder andererseits überprüfen, ob ein bestimmtes Element ein bestimmtes Ziel erreicht (z. B. ob ein Informationselement korrekte

---

und relevante Informationen auf unmissverständlich Art und Weise vermittelt).

Eine frühzeitige Klärung des Evaluationsgegenstandes ist vor allem mit Blick auf die Machbarkeit der Evaluation wichtig. Zudem sollte der Evaluationsgegenstand dokumentiert werden: „Sowohl das Konzept des Evaluationsgegenstands als auch seine Umsetzung sollen genau und umfassend beschrieben und dokumentiert werden.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 27, G 1 Beschreibung des Evaluationsgegenstandes)

### **Ist die Evaluation unter den Rahmenbedingungen realistisch machbar?**

Bei der Frage der Machbarkeit sind u.a. die in der Arbeitsgruppe vorhandene Expertise, die zur Verfügung stehenden finanziellen Ressourcen und die Komplexität des Evaluationsgegenstands und der Evaluationsfragestellungen relevant.

#### **Wichtig**

Um Machbarkeit, Transparenz und Akzeptanz des Evaluationsvorhabens inklusive der Nutzung der Evaluationsergebnisse zu gewährleisten, sollten frühzeitig folgende Aspekte berücksichtigt werden:

- a. Klärung des Evaluationszwecks
- b. Überprüfung und ggf. Stärkung der Glaubwürdigkeit der Evaluation
- c. Identifizierung der relevanten Interessensträgerinnen und -träger
- d. Identifikation und Dokumentation des Evaluationsgegenstandes
- e. Einschätzung der Machbarkeit des Vorhabens unter Berücksichtigung von Gegenstand, Zweck und zur Verfügung stehender Ressourcen

### 9.2.2 Wahl eines Evaluationsansatzes – Vom Zweck zur Fragestellung

Der Evaluationsansatz ergibt sich aus dem Zweck der Evaluation (Döring & Bortz, 2016; s. Abschn. 9.2.1). Dabei kann zwischen summativen und formativen sowie zwischen prozessbegleitenden und vergleichenden Evaluationsansätzen unterschieden werden. Welcher Ansatz gewählt wird, ist zum einen relevant, um sich über die Bedeutung der Rolle der für die Evaluation zuständigen Personen klarzuwerden (so ist die Unabhängigkeit und damit Glaubwürdigkeit bei der summativen Evaluation von höchster Relevanz, bei der formativen Evaluation aber weniger relevant). Zum anderen sind mit der Wahl eines Evaluationsansatzes das methodische Vorgehen (z. B. Untersuchungsdesigns und Erfassungsmethoden) und die Form der Berichtslegung (z. B. kontinuierlich und informell vs. einmalig und formell) eng verknüpft. Die Festlegung auf einen Evaluationsansatz kann also methodische Entscheidungen bahnen und vereinfachen.

#### **Summative vs. formative Evaluation**

„When the cook tastes the soup, that's formative. When the guests taste the soup, that's summative.“, so lautet Robert Stakes Metapher zu den beiden wichtigsten Ansätzen in der Evaluationsforschung (zitiert nach Hattie, 2013, S. 164). Übertragen auf den OSA-Kontext zeigt diese Analogie, dass es sich um eine summative (ergebnisorientierte) Evaluation handelt, wenn überprüft wird, ob das OSA die erwarteten Ziele erreicht (z. B. eine gesteigerte Passung zwischen Studierenden und Anforderungen im Studiengang). Die formative Evaluation hingegen dient den OSA-Entwicklerinnen und -Entwicklern als Unterstützung bei der Gestaltung und Optimierung des OSA und adressiert Fragen der Zufriedenheit, Nutzerfreundlichkeit, Akzeptanz aber auch Fragen dazu, welche OSA-Elemente besonders zur Wirkung beitragen.

Bei der summativen Evaluation werden schwerpunktmäßig quantitative Methoden und ‚starke‘ (d. h. intern valide) Untersuchungsdesigns eingesetzt (s. Abschn. 9.4), um möglichst belastbare Evaluationsergebnisse zu erhalten. Bei der formativen Evaluation kommen häufig qualitative Methoden zum Einsatz, um neuartige und vielfältige Informationen darüber zu erhalten, wie ein OSA optimiert werden kann (z. B. die Methode des lauten Denkens, um Herauszufinden, wie die

---

Navigation im OSA nutzerfreundlicher gestaltet werden kann). Bei der summativen Evaluation wird erst am Ende des Evaluationsprozesses berichtet. Bei der formativen Evaluation dagegen ist es üblich und sinnvoll, dass die mit der Evaluation betrauten Personen in ständigem Austausch mit den OSA-Entwicklerinnen und -Entwicklern bezüglich der aktuell vorliegenden Ergebnisse stehen. Nicht zuletzt unterscheiden sich die beiden Ansätze auch in ihren Zielen und den herangezogenen Kriterien (s. Abschn. 9.2.3 und 9.3). Die Box zu den „Beispielfragestellungen für die Evaluation von Online Self-Assessments“ zeigt exemplarisch einige Fragestellungen der summativen und formativen Evaluation und der prozessbegleitenden Evaluation.

### **Beispielfragestellungen für die Evaluation von Online Self-Assessments**

#### Summative Evaluation

1. Wirkt sich die Teilnahme an einem OSA auf die Passung der Studienanfängerinnen und -anfänger zu den Anforderungen des Studiengangs aus?
2. Führt die Einführung eines OSA für einen Studiengang zu einer verringerten Studienabbruchrate nach dem ersten Fachsemester?
3. Ist die Studienzufriedenheit von Studierenden, die sich aufgrund der Teilnahme an einem OSA für den Studiengang entschieden haben, größer?

#### Formative Evaluation

4. Wird die anvisierte Zielgruppe ausreichend gut erreicht?
5. Ist die Gestaltung des OSA nutzerfreundlich?
6. Läuft das OSA auf unterschiedlichen Internetbrowsern störungsfrei?

#### Prozessevaluation

1. Verändert sich die Zusammensetzung der Gruppe der Nutzerinnen und Nutzern von OSA über die Zeit (z. B. im Hinblick auf ihre Vorkenntnisse)?
2. Welche Auswirkungen haben hochschulpolitische Entscheidungen (z. B. über Hochschulzulassung) auf die Nützlichkeit von OSA?
3. Wie verändert sich die Nutzerfreundlichkeit mit längerer „Lebensdauer“ des OSA?

#### Vergleichende Evaluation

1. Welchem der OSA gelingt es besser, die Zielgruppe zu erreichen und warum?
2. Welches OSA hat eine transparentere Nutzerführung?
3. Wie unterscheidet sich die Vorhersagekraft der eingesetzten diagnostischen Elemente für vergleichbare Studienerfolgskriterien bei zwei OSA zu demselben Studiengang an unterschiedlichen Standorten?

### **Prozessevaluation**

Die Prozessevaluation erfolgt nach Abschluss der Entwicklung eines OSA. Sie überprüft in regelmäßigen Abständen die Umsetzung des OSA im Anwendungsalltag. Sie wird außerdem als Implementationskontrolle begleitend zu einer summativen Evaluation eingesetzt, um die so genannte ‚treatment fidelity‘ (Ausführungsintegrität) zu überprüfen. Eine systematische Prozessevaluation kann außerdem dazu beitragen, die Nachhaltigkeit eines OSA zu sichern. Wenn rechtzeitig erkannt wird, dass bestimmte Veränderungen die Wirksamkeit der ursprünglichen OSA-Konzeption beeinträchtigen könnten, kann entsprechend reagiert und eine Anpassung des OSA vorgenommen werden. Solche Änderungen können z. B. eine sich ändernde Zielgruppe oder Änderungen der technischen Voraussetzungen sein. Ein prototypisches Beispiel für eine Prozessevaluation ist die Nutzerbefragung nach Absolvieren eines OSA (s. Box „Praxisbeispiel: Fragen zur begleitenden Prozessevaluation von OSA“)

### **Praxisbeispiel: Fragen zur begleitenden Prozessevaluation von OSA**

Die folgenden Fragen wurden vom Netzwerk Online Self-Assessment erarbeitet um zentrale Aspekte der OSA-Nutzung direkt im Rahmen des OSA einschätzen zu lassen.

#### **Evaluationsfragen zur Erfassung im Anschluss an die Bearbeitung des OSA**

Diese Fragen werden einmalig im Anschluss an die Teilnahme am OSA erhoben. (Antwortskala: 0 = „stimmt überhaupt nicht“ bis 5 = „stimmt genau“)

- |                            |   |
|----------------------------|---|
| <b>Informiertheit</b>      | - Durch dieses Online-Self-Assessment kann ich meine Stärken und Schwächen besser einschätzen.                        |
|                            | - Durch dieses Online-Self-Assessment kann ich die Anforderungen des Studiengangs X besser einschätzen.               |
|                            | - Durch dieses Online-Self-Assessment sind meine Erwartungen an ein Studium der X in Y realistischer geworden.        |
| <b>Passung und Eignung</b> | - Durch dieses Online-Self-Assessment kann ich besser einschätzen, ob ein Studium der X zu meinen Interessen passt.   |
|                            | - Durch dieses Online-Self-Assessment kann ich besser einschätzen, ob Studiengang X in Y zu meinen Fähigkeiten passt. |
| <b>Nutzen</b>              | - Das Online-Self-Assessment ist mir eine Hilfe bei der Entscheidungsfindung.   |
| <b>Akzeptanz</b>           | - Ich werde dieses Online-Self-Assessment anderen weiterempfehlen, die sich für dieses Studium interessieren.         |

#### **Evaluationsfragen zur Erfassung vor und nach der Bearbeitung des OSA**

Diese Fragen werden vor und nach der Teilnahme am OSA erhoben. (Antwortskala: 1 = „sehr sicher, dass ich nicht studieren will“, 3 = „noch unsicher“ bis 5 = „sehr sicher, dass ich es studieren möchte“)

- |                                    |   |
|------------------------------------|---|
| <b>Studienwahl-<br/>sicherheit</b> | - Wie sicher sind Sie, dass Sie Studiengang X studieren möchten?        |
|                                    | - Wie sicher sind Sie, dass Sie an der Universität Y studieren möchten? |

### **Vergleichende Evaluation**

Unter den Begriff der vergleichenden Evaluation fällt jede Evaluation, die zwei oder mehr OSA vergleicht. Ein Vergleich zweier unabhängiger OSA bezüglich deren Wirksamkeit im Rahmen der summativen Evaluation erscheint sinnvoll und möglich, wenn die OSA dieselben Ziele verfolgen und vergleichbare Funktionen und Wirkmechanismen adressieren. Diese Vergleichbarkeit der Ziele und Funktionen ist wichtig, denn ein OSA, welches beispielsweise zum Ziel hat, die Eignung der Studienbewerberinnen und -bewerber für einen bestimmten Studiengang zu erhöhen, kann nicht mit einem OSA verglichen werden, welches der Studienorientierung dient. Ein Vergleich zweier OSA im Hinblick auf Fragen der Gestaltung (z. B. nutzerfreundliche Gestaltung, technische Stabilität, u. ä.) wie sie bei der formativen oder auch der Prozessevaluation untersucht werden, erscheint dann sinnvoll, wenn das Ziel die Identifikation einer ‚optimalen‘ Gestaltung ist, d. h. aus den Gestaltungsformen anderer OSAs zur Optimierung der Gestaltung gelernt werden soll. Voraussetzung ist auch hier, dass die anvisierten Funktionen und Wirkmechanismen vergleichbar sind. Bislang existieren zwar einige Studien zu einer vergleichenden Bewertung von OSA (Bildung und Begabung, 2021; Brunner, 2017; Brunner, Kretschmer & Maschwitz, 2017; Gollub & Meyer-Guckel, 2014). Deren Aussagekraft ist jedoch unklar, da die oben genannten Anforderungen an vergleichende Studien (d. h. vergleichbare Ziele und Funktionen, vergleichbare Wirkmechanismen, etc.) nur eingeschränkt oder gar nicht berücksichtigt wurden.

---

### **Stolpersteine bei der vergleichenden summativen Evaluation von OSA**

Selbst bei vergleichbaren Zielen und Funktionen von verschiedenen OSA-Plattformen bleibt unklar, welche Bedeutung der Standort für die Vergleichbarkeit von Evaluationsergebnissen hat. So ist ein OSA in der Regel an den Standort gebunden und setzt dessen Spezifika um (z. B. durch die Normierung der diagnostischen Elemente an aktuell erfolgreichen Studierenden, s. Kap. 5). Passung und Eignung an einem Standort entsprechen dabei nicht unbedingt der Passung und Eignung bei einem anderen. Unterschiedliche Standorte haben unterschiedliche (ggf. auch an das Bundesland gebundene) Zulassungsverfahren, weswegen Studieninteressierte unterschiedliche Voraussetzungen mitbringen. Auch basiert das Studienerfolgskriterium der Fach-, Modul- oder Abschlussnoten in verschiedenen Studiengängen auf verschiedenen Notengebungsverfahren. Die Frage der Standortspezifität ist jenseits ihrer Bedeutung für die Interpretation von Evaluationsergebnissen jedoch auch (hochschul-)politisch relevant, wenn es um die politischen Entscheidungen bzgl. der Gestaltung der OSA-Landschaft in Bundesländern, Hochschulverbänden etc. geht. Empirische Forschung dazu steht aus.

### 9.2.3 Bestimmung der Evaluationskriterien

Welche Kriterien relevant sind, hängt von den Zielen des OSA, dem Evaluationsansatz und den daraus abgeleiteten Evaluationsfragestellungen ab. Kriterien sollten dabei zentrale Aspekte der Ziele repräsentieren, sie sollten außerdem geeignet sein, die Zielerreichung zu überprüfen (Döring & Bortz, 2016). Die Festlegung der Kriterien ist ein zentraler Bestandteil eines transparenten Evaluationsprozesses (Balzer, 2005; DeGEval – Gesellschaft für Evaluation e.V., 2017).

Gemäß der Evaluationsstandards sollten zur Gewährleistung einer umfassenden und fairen Prüfung (DeGEval – Gesellschaft für Evaluation e.V., 2017, F 3 Umfassende und faire Prüfung), der Berücksichtigung der Informationsbedürfnisse der Beteiligten und Betroffenen (DeGEval – Gesellschaft für Evaluation e.V., 2017, N 4 Auswahl und Umfang der Informationen) sowie der unparteilichen Durchführung und Berichterstattung (DeGEval – Gesellschaft für Evaluation e.V., 2017, F 4 Unparteiische Durchführung und Berichterstattung) idealerweise alle Interessensträgerinnen und -träger in die Ableitung der Kriterien einbezogen werden, da die mit dem OSA verbundenen Ziele für unterschiedliche Interessensgruppen verschieden, wenn nicht sogar gegensätzlich, sein können. So kann die Berücksichtigung aller Interessensträgerinnen und -träger und ihrer Ziele die Gefahr der missbräuchlichen Verwendung der Evaluation verringern. Zur Bestimmung von Kriterien für die OSA-Evaluation lassen sich abgeleitet aus Döring und Bortz (2016) drei Wege identifizieren.

#### **Top-down- oder deduktiver Ansatz (Konzeptions-gebundene Kriterien)**

Idealerweise können Kriterien aus der OSA-Konzeption entnommen werden, diese sollte Informationen darüber enthalten, mit welchen Zielen das OSA entwickelt wurde (z. B. Verbesserung des Studieneinstiegs im Studiengang Chemie durch eine bessere Eignung der Studienanfängerinnen und -anfängern; s. Kap. 2 und s. Kap. 3). Konzeptionsgebundene Ziele und die passenden Kriterien geben in der Regel darüber Auskunft, was durch das OSA erreicht werden soll. Sie stellen somit eine zentrale Grundlage einer summativen Evaluation dar. In Abschn. 9.3 wird die Frage geeigneter Kriterien zur summativen Evaluation näher beleuchtet.

#### **Bottom-up- oder induktiver Ansatz (Interessensgruppen-gebundene Kriterien)**

Alle relevanten Interessensgruppen sollten in die Bestimmung relevanter und geeigneter Evaluationskriterien eingebunden werden. Unter der Bedingung, dass alle Interessensgruppen berücksichtigt wurden und der Prozess gut dokumentiert wurde, ist es möglich, die Ergebnisse der Anforderungsanalyse aus der OSA-Entwicklung einzubeziehen (s. Kap. 3). Unerwünschte Nebenwirkungen (d. h. Ängste davor, Studierende abzuschrecken), die Sicht potentieller Nutzerinnen und Nutzer und die der OSA-Arbeitsgruppe selbst sollten dabei nicht vergessen

---

werden, da sie häufig als Kriterien für eine formative Evaluation genutzt werden können. Die Box „Beispiele für Kriterien zur formativen Evaluation von OSA“ fasst prototypische Beispiele für formative Evaluationskriterien in sechs übergeordneten Kategorien zusammen.

Bei der Interessensgruppen-gebundenen Kriterienbestimmung stehen zu Beginn häufig sehr vage formulierte Fragestellungen (z. B. „Bringt das OSA was?“; „Kommen die Studieninteressierten mit dem OSA klar?“). Diese müssen durch die für die Evaluation zuständigen Personen konkretisiert und dabei in Ziele (z. B. Steigerung des Studienerfolgs, Akzeptanz des OSA), Kriterien (z. B. Studiengeschwindigkeit, adressatenorientierte Sprache) und passende und empirisch beobachtbare Indikatoren (z. B. ECTS in Relation zu der Zahl der Fachsemester, Zahl der genutzten Fremdwörter) „übersetzt“ werden. Nur wenn für die Kriterien messbare Indikatoren festgelegt werden, kann die Erreichung der Kriterien auch empirisch überprüft werden. Die in Kap. 5 beschriebenen Grundsätze zur Wahl oder Entwicklung von diagnostischen Instrumenten können dabei auch auf diagnostische Elemente übertragen werden, welche summative und formative Evaluationskriterien erfassen sollen.

#### **Beispiele für Kriterien zur formativen Evaluation von OSA**

**Akzeptanz:** Zufriedenheit mit der Teilnahme am OSA, Wahrgenommene Nützlichkeit bei der Studienwahlentscheidung, Ernsthaftigkeit und Vollständigkeit der Bearbeitung des OSA (Compliance) etc.

**Nutzerfreundlichkeit (Usability):** Geeignete Ansprache der Zielgruppe vor und während der Durchführung des OSA, Transparenz der Nutzerführung, Verfügbarkeit von technischem Support etc.

**Ökonomie:** Ausgewogenes Verhältnis von Kosten und Nutzen, angemessene Bearbeitungsdauer etc.

**Nachhaltigkeit:** Auf langfristige Nutzbarkeit ausgerichtete Gestaltung des OSA, ausreichende personelle und finanzielle Ressourcen, kontinuierliche Prozessevaluation etc.

**Qualität der Konzeption:** Beteiligtenorientierung, Berücksichtigung der gesellschaftlichen und (hochschul-)politischen Rahmenbedingungen, Dokumentation der Gütekriterien der diagnostischen Instrumente etc.

**Qualität der Implementation:** Auffindbarkeit durch Studieninteressierte, Technische Stabilität, Technische Flexibilität

#### **Normorientiertes Vorgehen (normative Kriterien, auch Qualitätsstandards)**

Dieser Ansatz basiert auf der Annahme, dass es Gütekriterien oder allgemeingültige Standards gibt, die universell für eine zu evaluierende Maßnahme gelten und für die Beurteilung der Qualität dieser Maßnahme angewendet werden können. Beispiele aus anderen Kontexten sind Hygienestandards bei der Produktion oder Verarbeitung von Lebensmitteln oder auch generell DIN- und ISO-Normen. Für OSA existieren bislang keine allgemeingültigen Normen oder Qualitätsstandards. Die DIN 33430 (DIN, 2016; s. Kap. 5), welche Anforderungen an die berufsbezogene Eignungsdiagnostik beschreibt, ist in Teilen auf die diagnostischen Elemente in OSA übertragbar. Schmitt und Schmidt-Atzert (2019) liefern einen ersten Ansatz in Richtung OSA-spezifischer Qualitätsstandards. In den Kapiteln dieses Buches sind diese genannten Ansätze eingebunden und ergänzt worden, eine Zusammenstellung findet sich in s. Kap. 10. Qualitätsstandards sind in der Regel für die formative Evaluation oder die Prozessevaluation relevant.

Die genannten Ansätze sollten in Ergänzung miteinander genutzt werden und geeignete und passende Kriterien wohlbedacht ausgewählt werden. Ziele und Kriterien können dabei miteinander in Konflikt stehen: Wählt man Kriterien deduktiv, besteht die Möglichkeit, berechnete Interessen der Interessensträgerinnen und -träger zu vernachlässigen, die man nur über das

---

induktive Vorgehen identifizieren kann. Konzeptionsgebundene Kriterien können beispielsweise auf geringe Akzeptanz bei den Interessensträgerinnen und -trägern stoßen. So kann eine durch die Konzeption des OSA angestrebte Steigerung der Eignung und Passung von Studienanfängerinnen und -anfängern mit der Befürchtung von Hochschulleitungen und Studiengangverantwortlichen im Konflikt stehen, dass Studieninteressierte durch die Teilnahme am OSA von einer Bewerbung auf einen Studienplatz abgeschreckt werden. Ebenso sollten fairerweise keine Kriterien eingesetzt werden, die gemäß OSA-Konzeption gar nicht erreicht werden sollten. Bessere Studiennoten der OSA-Teilnehmerinnen und -Teilnehmern sollten demnach kein Evaluationskriterium sein, wenn es nicht auch das Ziel des OSA war, dass eine Teilnahme zu besseren Studiennoten führt.

Derartige Dilemmata bei der Bestimmung der Kriterien sollten idealerweise antizipiert werden, um sowohl fachlich-methodisch als auch argumentativ vorbereitet zu sein: „Evaluationen sollen so geplant und durchgeführt werden, dass eine möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 22, D 2 Diplomatisches Vorgehen).

#### 9.2.4 Methodische Planung, Durchführung und Auswertung von Evaluationsstudien

Bereits mit der Festlegung auf einen Evaluationsansatz (summativ vs. formativ) sind die Weichen für die methodische Ausgestaltung einer Evaluationsstudie gestellt. Die Planung, Durchführung und Auswertung entspricht dann anderen empirischen Studien ohne evaluatorischen Charakter. Grundsätzlich sollen „Erhebungsverfahren und Datenquellen [...] so gewählt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit [=Validität] bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien der empirischen Forschung orientieren.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 29, G 5 Valide und reliable Informationen).

Eine erschöpfende Darstellung aller relevanten methodischen, diagnostischen und statistischen Grundlagen und der relevanten Gütekriterien ist in diesem Kapitel weder möglich noch sinnvoll. Für den interessierten Leser verweisen wir auf die einschlägige Lehrbuchliteratur (Beauducel & Leue, 2014 für Psychologische Diagnostik; Döring & Bortz, 2016 für Untersuchungsplanung, Forschungsmethoden und Evaluation; Eid, Gollwitzer & Schmitt, 2017 für Statistik und Forschungsmethoden). Die vorliegende Darstellung beschränkt sich daher auf einen Überblick über die methodischen und statistischen Entscheidungen, die im Rahmen der Planung, Durchführung und Auswertung von Evaluationsstudien getroffen werden müssen. Die Box „Leitfragen zu Planung, Durchführung und Auswertung von empirischen Studien“ listet – wenn auch ohne Anspruch auf Vollständigkeit – wichtige Leitfragen auf. Diese dienen der Strukturierung des Vorgehens und weisen auf die wichtigsten methodischen Fragen hin, die im Rahmen der methodischen Ausgestaltung einer Evaluationsstudie zu klären sind.

### Leitfragen zu Planung, Durchführung und Auswertung von empirischen Studien

<b>Forschungsdesign</b>	<ul style="list-style-type: none"><li>- Welches Untersuchungsdesign wird angewendet?</li><li>- Wie wird die Güte des Designs und dessen Aussagekraft beurteilt und sichergestellt?</li></ul>
<b>Stichprobe</b>	<ul style="list-style-type: none"><li>- Wie wird die Stichprobe rekrutiert?</li><li>- Wie groß muss die Stichprobe sein, um eine ausreichende statistische Power zu erreichen?</li></ul>
<b>Untersuchungsmaterialien</b>	<ul style="list-style-type: none"><li>- Welche Erhebungsinstrumente werden eingesetzt?</li><li>- Ist die diagnostische Güte der Instrumente ausreichend?</li></ul>
<b>Durchführung</b>	<ul style="list-style-type: none"><li>- Wird die Studie online oder im Paper-und-Bleistift Format durchgeführt?</li><li>- Wie wird sichergestellt, dass Bearbeitungszeiten eingehalten werden?</li></ul>
<b>Statistische Auswertung</b>	<ul style="list-style-type: none"><li>- Wie werden die Daten aufbereitet und bereinigt?</li><li>- Sind die Voraussetzungen für die geplanten statistischen Verfahren erfüllt?</li></ul>

In Kap. 10 werden die wichtigsten, mit den einzelnen methodischen Fragen verbundenen Gütekriterien (z. B. Gütekriterien von Untersuchungsplänen) eingeordnet und beschrieben. Sie sollten zur Sicherstellung der Aussagekraft, Belastbarkeit und Nützlichkeit von summativen und formativen Evaluationsstudien berücksichtigt werden. Aufgrund der besonderen Relevanz wird die Frage, welche versuchsplanerischen Überlegungen bei der Wahl eines adäquaten Forschungsdesigns zur Untersuchung der Wirksamkeit von OSA maßgeblich sind, in Abschn. 9.4 näher beleuchtet.

#### 9.2.5 Berichtstellung: Interpretation der Ergebnisse und Meta-Evaluation

Der Schlusspunkt der Evaluation stellt die Interpretation der Ergebnisse und die Meta-Evaluation im Rahmen der Berichtstellung dar. Beides sollte noch durch die für die Evaluation Zuständigen so vorgenommen werden, dass die Ergebnisse zum ursprünglichen Zweck der Evaluation optimal genutzt werden können.

Abhängig vom Zweck der Evaluation sollten bei der Interpretation inhaltliche Antworten auf die gestellten Evaluationsfragestellungen gegeben und Implikationen für Umsetzung der Ergebnisse (z. B. für die OSA-Entwicklung) abgeleitet werden. Dabei sollen „die in einer Evaluation getroffenen wertenden Aussagen [...] auf expliziten Kriterien und Zielwerten basieren. Schlussfolgerungen sollen ausdrücklich und auf Grundlage der erhobenen und analysierten Daten begründet werden, damit sie nachvollzogen und beurteilt werden können.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 30, G 8 Begründete Bewertungen und Schlussfolgerungen). Vermutungen und Spekulationen über das hinaus, was die empirischen Evidenzen erlauben, sollten dabei unbedingt vermieden werden.

Bei der Meta-Evaluation (d. h. Evaluation der Evaluation) sollen „die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen [...] systematisch auf Fehler geprüft werden.“ (DeGEval – Gesellschaft für Evaluation e.V., 2017, S. 29, G 6 Systematische Fehlerprüfung) und die hierbei identifizierten Grenzen der Aussagekraft der Ergebnisse ohne Einschränkungen beschrieben werden. Beispiele für prototypische Limitation

---

sind die Stichprobe (Größe oder Zusammensetzung), inadäquate oder eingeschränkt valide Untersuchungsdesigns, ad-hoc entwickelte Erhebungsinstrumenten, die nicht ausreichend auf ihre diagnostische Güte hin überprüft wurden, fehlende oder unvollständige Erfassung von Störvariablen.

Die Berichtstellung verfolgt das Ziel, die Glaubwürdigkeit des evaluatorischen Vorgehens zu stärken, Nutzende der Ergebnisse mit nötigen Informationen auszustatten, um die Evaluation des OSA methodenkritisch beurteilen zu können und somit den Nutzen und die Nutzbarkeit der Evaluationsergebnisse zu optimieren. So sollte man sich auch und gerade bei mündlichen Berichten nicht zu Antworten auf Fragen verleiten lassen, die auf Basis der Evaluationsergebnisse nicht zulässig sind und die im Rahmen der Evaluation nicht beantwortet werden konnten oder sollten.

### **9.3 Kriterien zur Evaluation der Wirksamkeit von OSA**

In Kap. 3 (Anforderungsanalyse) werden wichtige Kriterien, die mit den Zielen der OSA-Entwicklung verbunden sind, umfassend beschrieben. Diese sind v. a. aus für OSA einschlägigen Literaturquellen entnommen (Ott, Ulfert & Kersting, 2017; Thiele & Kauffeld, 2020; Zimmerhofer, Heukamp & Hornke, 2006) und weitestgehend identisch mit klassischen Studienerfolgskriterien wie Studienabbruch oder Studiennoten, man findet sie zudem in der eignungsdiagnostischen Forschung zur Vorhersage von Studienerfolg (Trapmann, Hell, Hirn & Schuler, 2007; Trapmann, Hell, Weigand & Schuler, 2007). Betrachtet man OSA vor allem als eine diagnostische Plattform, auf der Studieninteressierte eine Rückmeldung zu ihren erfolgsrelevanten Voraussetzungen erhalten, erscheinen die klassischen Studienerfolgskriterien zunächst geeignet für eine Evaluation der OSA-Wirksamkeit.

Wir betrachten OSA hingegen als komplexe Intervention, deren Wirkung über die Vorhersagekraft der diagnostischen Elemente hinausgehen kann. Die angenommenen OSA-Wirkmechanismen Selbstreflexion und Selbstselektion, Fokus auf Passung und Eigenverantwortung (s. Kap. 1) sollen dazu beitragen, dass der Prozess der Studienwahlentscheidung durch die Bearbeitung des OSA verbessert wird (Milbradt, Zettler, Putz, Heukamp & Hornke, 2008; Thiele & Kauffeld, 2020). Dies beinhaltet idealerweise eine höhere Qualität und Strukturiertheit der genutzten Informationen (z.B. zu den Anforderungen im Studium oder den studienerelevanten Personenmerkmalen), eine größere Verarbeitungstiefe und kritischere Reflektion der Informationen durch die Studieninteressierten aufgrund einer durch die Plattform angeleiteten Informationsverarbeitung. Gemäß dieser Annahmen sollte sich die Bearbeitung eines OSA zunächst auf die Studienwahlentscheidung selbst auswirken (Heukamp, Putz, Milbradt & Hornke, 2009).

Da das OSA sowohl die personenseitigen Voraussetzungen als auch die späteren Anforderungen darstellt und miteinander in Bezug setzt, sollte die Bearbeitung des OSA vermittelt über die Mechanismen der Selbstreflexion und Selbstselektion zu einer besseren realistischeren Einschätzung der eigenen Passung und Eignung beitragen, so dass sich diejenigen mit einer geringeren Passung und Eignung eher gegen die Aufnahme dieses bestimmten Studiums entscheiden. Studienanfängerinnen und -anfänger, die das OSA absolviert haben und sich dann für einen Studiengang entschieden haben, sollten daher im Mittel eine höhere Passung zu und Eignung für den gewählten Studiengang aufweisen als die Studienanfängerinnen und -anfänger, die das OSA nicht absolviert haben (Thiele & Kauffeld, 2020). Durch die bessere Passung sollte zudem die subjektive Bewertung des Studiums beispielsweise in Form von Studienzufriedenheit, der erlebten Passung und Eignung, der Motivation zum oder das Stresserleben durch das Studium besser ausfallen (Thiele & Kauffeld, 2020).

Fairerweise muss daher die Frage gestellt werden, ob oder in welchem Ausmaß die eher klassischen Studienerfolgskriterien geeignet sind, die Wirksamkeit eines OSA zu evaluieren. Studienerfolgskriterien können häufig erst lange nach Bearbeitung des OSA beobachtet und erfasst werden und stehen somit nur in einem indirekten Zusammenhang mit den OSA-Wirkmechanismen. Darüber hinaus werden die klassischen Studienerfolgskriterien von

zahlreichen weiteren Faktoren beeinflusst (z. B. weitere Rahmenbedingungen oder zwischenzeitliche Geschehnisse während des Studiums).

Abb. 9.1 veranschaulicht diese Überlegungen. Die Wirkungen, die direkt aus der OSA-Teilnahme resultieren, werden hier als erfolgsvermittelnde Faktoren dargestellt, die in einem direkten Zusammenhang mit den Wirkmechanismen des OSA stehen. Erst wenn das OSA seine Wirkung in Form dieser erfolgsvermittelnden Faktoren zeigt, kann daraus ein gesteigerter Studienerfolg (objektiv und subjektiv) resultieren. Dass eine gesteigerte Passung und Eignung und auch eine bessere Bewertung des Studiengangs relevante Determinanten für Studienerfolg darstellen, konnte bereits gezeigt werden (Kubinger, Frebort & Müller, 2012). Zudem gibt es erste Hinweise, dass OSA, die im Laufe des Studiums eingesetzt werden, die Selbstreflexion (Meta-Kognition) fördern und so zu besseren Leistungen im Studium beitragen können (Ibabe & Jauregizar, 2010).

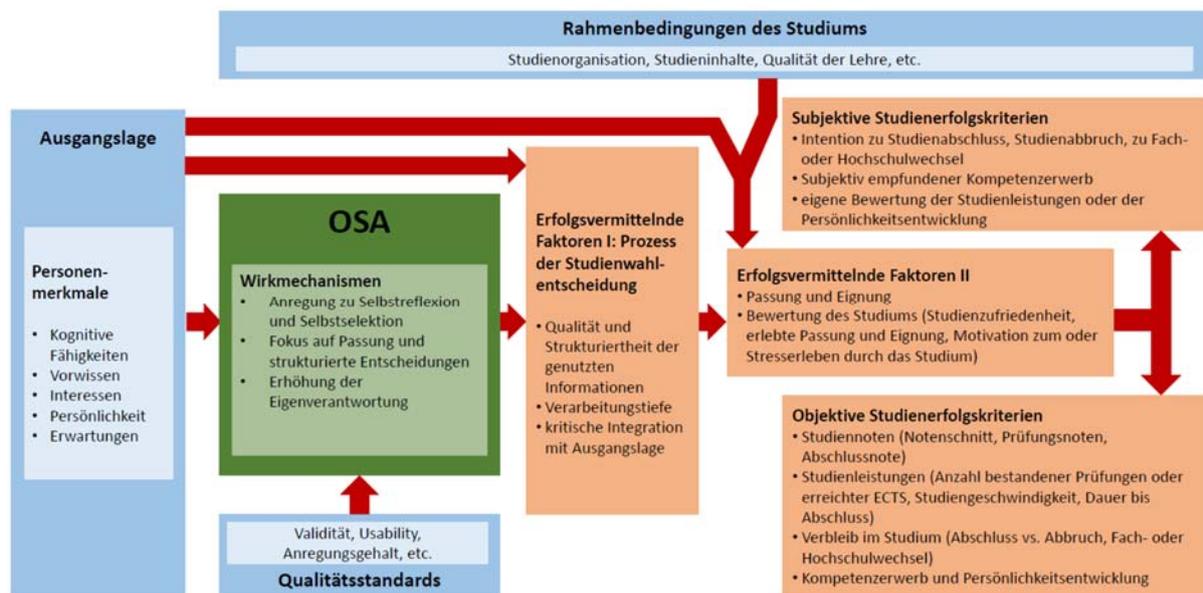


Abb. 9.1 Kriterien zur summativen Evaluation von OSA entlang eines erweiterten Rahmenmodells

**Fehler! Verweisquelle konnte nicht gefunden werden.** zeigt die individuelle Perspektive der Studieninteressierten und späteren Studierenden, da bei ihnen die Wirkmechanismen zum Tragen kommen sollen. Aus der Perspektive institutioneller Interessensgruppen können aber ebenfalls Fragestellungen an die Evaluation herangetragen werden. Aus den angenommenen Wirkmechanismen des OSA sind diese Fragestellungen und v. a. die zugehörigen Evaluationskriterien jedoch nicht direkt ableitbar. Um also geeignete Evaluationskriterien auf institutioneller Ebene zu bestimmen, können die Wirkungen auf individueller Ebene als Ausgangspunkt betrachtet werden (z. B. eine verbesserte Studienwahlentscheidung oder eine höhere Passung von Studienanfängerinnen und -anfängern). Tritt eine solche Wirkung ein, ist anzunehmen, dass diese sich im Folgenden auch auf die Ebene des Studiengangs oder der Hochschule auswirkt (z. B. ein verringerter Beratungsaufwand oder eine homogenere Zusammensetzung der Erstsemester). Tab. 9.2 zeigt mögliche Fragestellungen aus der Perspektive von Studiengang oder Hochschule und nennt die zugehörigen Kriterien auf Ebene des Individuums, aus denen die Kriterien auf Ebene von Studiengang und Hochschule abgeleitet werden können.

Tab. 9.2 Institutionelle Evaluationsfragestellungen und damit im Zusammenhang stehende Kriterien

Fragestellung	Kriterien im Wirkzusammenhang	
	Individuum	Studiengang/Hochschule
Wie entwickelt sich die Bewerberlage nach der Einführung eines OSA?	Verbesserte Studienwahlentscheidung; höhere Passung und Eignung	Geringere Bewerberzahlen und effizientere Auswahl; Anteil der Geeigneten unter allen Bewerberinnen und Bewerbern; reduzierter Beratungsaufwand
Wie verändert sich der Beratungsaufwand bei der (Fach-)Studienberatung nach Einführung des OSA?	Höhere Passung und Eignung, bessere Bewertung des Studiums	Zahl und Qualität der Beratungsanfragen
Wie entwickelt sich das Interesse an Brückenkursen?	Höhere Passung und Eignung, bessere Bewertung des Studiums	Zahl der Teilnehmerinnen und Teilnehmer an Brückenkursen; Übereinstimmung von Anforderungen der Brückenkurse mit den Bedarfen der Teilnehmerinnen und Teilnehmer
Wie ändert sich die Zusammensetzung der Gruppe der Studienanfängerinnen und -anfänger?	Höhere Passung und Eignung, bessere Bewertung des Studiums	Homogenität der Erstsemester im Hinblick auf Interessen und Vorwissen, Anspruch an / Rückwirkung auf Lehrqualität und -organisation
Ändert sich die Wahrnehmung der Hochschule / des Studiums aus der Sicht von Arbeitgebern oder Absolventinnen und Absolventen?	Bessere Bewertung des Studiums, höherer objektiver und subjektiv wahrgenommener Kompetenzerwerb	Bindung an die Hochschule, Weiterempfehlung der Hochschule durch Absolventinnen und Absolventen oder Arbeitgeber
Lohnt sich die Einführung beziehungsweise die mittel- und langfristige Pflege von OSA-Plattformen finanziell?	Studienabbruch, Studien-geschwindigkeit	Verhältnis von Kosten der OSA-Entwicklung und dessen nachhaltiger Pflege zum finanziellen Nutzen (Hochschulfinanzierung auf Basis von Zahl der Absolventinnen und Absolventen in Regelstudienzeit)

## 9.4 Fragen der Untersuchungsplanung empirischer Studien zur summativen Evaluation von OSA

Im Rahmen der Untersuchungsplanung von empirischen Studien müssen verschiedene methodische Aspekte durchdacht und geklärt werden: das Forschungsdesign, die Stichprobe, die Untersuchungsmaterialien, die geplante Durchführung und die geplante statistische Auswertung (s. Abschn. 9.2.4). Aufgrund ihrer Bedeutung für die Aussagekraft summativer Evaluationsstudien greifen wir hier die Aspekte des Forschungsdesigns, der Stichprobe und der Untersuchungsmaterialien auf.

## 9.4.1 Wahl eines Forschungsdesigns zur Vermeidung oder Kontrolle von Konfundierung

Betrachtet man OSA als eine komplexe Intervention, kann die Beurteilung der Wirksamkeit eines OSA aus der Perspektive der Interventions- und Evaluationsforschung betrachtet werden. Dabei sind in Anlehnung an die Konzeption der fünf Datenboxen nach Wittmann (1990) drei Merkmalsbereiche von besonderer Bedeutung: Die Ausgangslage der Zielgruppe, die Wirkmechanismen des OSA und die Kriterien, an denen der Erfolg des OSA festgemacht wird. Das Rahmenmodell der OSA-Entwicklung (s. Kap. 1) greift bereits diese Merkmalsbereiche auf. Ausgangslage, Wirkmechanismen und Kriterien stehen dabei in einer wechselseitigen Beziehung. Um im Rahmen der Wirksamkeitsevaluation wie gefordert ‚starke‘ Forschungsdesigns mit hoher interner Validität zu nutzen (s. Abschn. 9.2.2), muss diese wechselseitige Beziehung der Merkmalsbereiche berücksichtigt werden.

Das Ausmaß, in dem interne Validität gegeben ist, wird bestimmt durch das Ausmaß, in dem mögliche Störvariablen kontrolliert werden können. Störvariablen können beispielsweise solche Personenmerkmale darstellen, die sowohl mit den OSA-Wirkmechanismen (z. B. der gesteigerten Eigenverantwortung) als auch den Kriterien zur Bewertung des OSA (z. B. einer verbesserten Studienwahlentscheidung) im Zusammenhang stehen. Ist dies der Fall, so spricht man von einer Konfundierung. Ein Beispiel für eine mögliche Konfundierung lässt sich anhand des Personenmerkmals Gewissenhaftigkeit beschreiben. So ist es denkbar, dass gewissenhaftere Personen eher an einem OSA teilnehmen, das OSA eher instruktionsgemäß und eigenverantwortlich bearbeiten und die Rückmeldung zu Passung und Eignung eher selbstkritisch reflektieren als weniger gewissenhafte Personen. Dies könnte dazu führen, dass ein OSA seine Wirkung bei gewissenhafteren Personen besser entfalten kann. Gleichzeitig steht Gewissenhaftigkeit nachgewiesenermaßen mit Studienerfolg (z. B. Lernerfolg in den Klausuren der ersten Semester) im Zusammenhang (Trapmann, Hell, Hirn, et al., 2007). Ein positives Ergebnis einer summativen Evaluation (z. B. bessere Studienleistungen von OSA-Teilnehmerinnen und -Teilnehmern) könnte so nicht eindeutig auf die Teilnahme am OSA zurückgeführt werden, sondern möglicherweise auf die höhere Gewissenhaftigkeit der OSA-Teilnehmerinnen und -Teilnehmer.

Abb. 9.2 veranschaulicht schematisch die Gefahr der Konfundierung. Personenmerkmale der Ausgangslage vor der Teilnahme am OSA können demnach sowohl die OSA-Wirkmechanismen (s. Kap. 1) als auch die möglichen Kriterien zur Beurteilung der OSA-Wirksamkeit (erfolgsvermittelnde Faktoren und Studienerfolgskriterien; s. Abb. 9.1) beeinflussen.

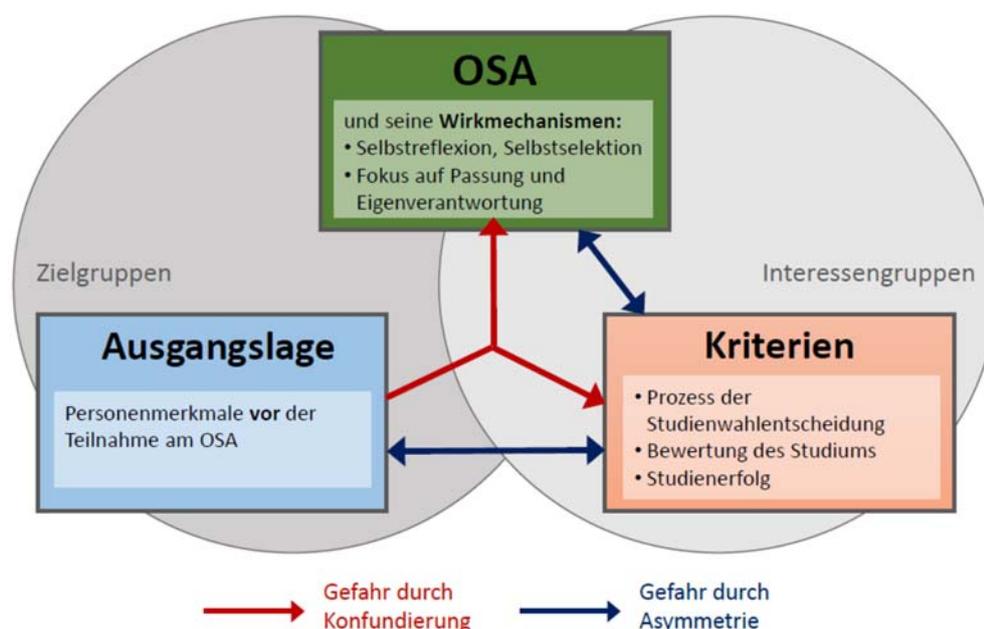


Abb. 9.2 Konfundierung und Asymmetrie als Gefahren für Evaluation der Wirksamkeit von OSA

---

Um eine mögliche Konfundierung zu vermeiden oder zu kontrollieren, sollte man zunächst mögliche Störvariablen identifizieren. Dazu ist es wichtig, sich die Intervention in Kombination mit ihrer anvisierten Wirkung aus theoretischer Sicht zu betrachten. Die Gefahr der Konfundierung beim OSA besteht zu drei Zeitpunkten: *Vor* der Teilnahme am OSA kann eine Konfundierung vorliegen, wenn mit den Kriterien zusammenhängende Personenmerkmale einen Einfluss darauf haben, wer überhaupt am OSA teilnimmt (z. B. das Personenmerkmal ‚Interesse an einem bestimmten Studienfach‘, welches sowohl die Wahrscheinlichkeit erhöht, dass jemand an einem bestimmten OSA teilnimmt als auch später die Bewertung des Studiums positiv beeinflusst). *Während* der OSA-Bearbeitung kann es zu einer Konfundierung kommen, wenn die Bearbeitung selbst und die Verarbeitung der Informationen durch erfolgsrelevante Personenmerkmale beeinflusst wird (s. Beispiel zur Gewissenhaftigkeit). *Nach* der Bearbeitung des OSA kann der Umgang mit der Rückmeldung und den Empfehlungen durch das OSA von erfolgsrelevanten Personenmerkmalen beeinflusst werden (z. B. persönliche Motive für ein Studium wie Geld verdienen vs. eine gute Ausbildung zu bekommen vs. in der Nähe von Freunden und Familie zu studieren), die die letztendliche Studienwahlentscheidung mit beeinflussen und ggf. auch den Studienerfolg beeinflussen können.

Auf Grundlage dieser theoretischen Überlegungen sollten bei der Wahl des Forschungsdesigns Maßnahmen zur Vermeidung oder Kontrolle der Konfundierung ergriffen werden. Zur *Vermeidung* von Konfundierung sind das idealerweise der Einsatz einer Kontrollgruppe (d. h. Personen, die das OSA nicht absolvieren) und die zufällige Zuweisung der Versuchspersonen zu Kontroll- oder Experimentalgruppe (d. h. die Gruppe von Personen, die das OSA absolvieren) (s. Beispielbox „Beispiel für ein Forschungsdesign zur summativen Evaluation der Wirksamkeit von OSA“). Die zufällige Zuweisung führt dazu, dass sich Störvariablen gleichmäßig und unsystematisch auf Personen in Experimental- und Kontrollgruppe verteilen und somit keinen Einfluss auf das Ergebnis der summativen Evaluation haben. Sollte eine zufällige Zuweisung nicht möglich sein (z. B. wenn die Versuchspersonen selbst entscheiden, ob sie am OSA teilnehmen oder nicht), so ist die *Kontrolle* von Störvariablen durch Parallelisierung (d. h. die Zusammenstellung von im Hinblick auf Störvariablen möglichst vergleichbare Experimental- und Kontrollgruppe) oder nachträgliche statistische Kontrolle notwendig. Dafür ist eine unabhängige, zeitlich vorgeschaltete Erfassung der Störvariablen absolut notwendig, um den Effekt der Störvariablen auf die Nutzung und Wirkung von OSA zu kontrollieren (s. Kap. 4 für eine Simulation, welche Auswirkung das Folgen der Empfehlung eines OSA auf die Vorhersagekraft der im OSA erfassten Instrumente hat). Für die Kontrolle von Störvariablen empfiehlt sich daher ein Prä-Post-Forschungsdesign, bei dem zusätzlich vor und nach der Teilnahme am OSA eine Befragung der Teilnehmenden erfolgt.

---

### **Beispiel für ein Forschungsdesign zur summativen Evaluation der Wirksamkeit von OSA**

Alle Schülerinnen und Schüler (SuS) des Abitur-Jahrgangs aller Gymnasien und Gesamtschulen in einer Universitätsstadt mit einer Voll-Universität absolvieren zu Beginn des letzten Schuljahres den Prätest. Bei allen SuS wird erhoben, ob die Absicht besteht zu studieren und welche Studiengänge in der engeren Auswahl sind (Studiengangpräferenz). Untersuchte Kriterien, die zum Prä- und Posttest erhoben werden, sind der Grad der Informiertheit über die interessierenden Studiengänge (erfasst über Selbstauskunft und anhand von Vorwissenstests zu Inhalten, Anforderungen und Organisation des Studiums), die subjektiv eingeschätzte Passung und Eignung zu den präferierten Studiengängen und die objektiv vorhandene Passung und Eignung auf Basis eines Abgleichs der als Störvariablen erfassten Prädiktoren für Studienerfolg mit der Normstichprobe aktuell erfolgreicher Studierender in den Studiengängen. Als potentielle Störvariablen werden kognitive Fähigkeiten, die Persönlichkeitsmerkmale Gewissenhaftigkeit, emotionale Stabilität und Offenheit für neue Erfahrungen, das Interessensprofil und der Bildungshintergrund der Eltern erfasst. Im Anschluss werden die SuS mit Studiengangpräferenz in einem von zwei bestimmten Studienfeldern (hier: Naturwissenschaft/Technik und Sozialwissenschaften) zufällig auf Experimentalgruppe (Teilnahme am OSA der Universität im Rahmen des Unterrichts) und Wartelisten-Kontrollgruppe (Teilnahme am OSA der Universität im Rahmen des Unterrichts sechs Wochen nach der Experimentalgruppe) aufgeteilt. Nach Absolvieren des OSA durch die EG werden die Kriterien (Vorwissen, Passung und Eignung jeweils subjektiv und objektiv) in EG und KG erneut erfasst. Außerdem werden wieder die Absicht zu studieren und die Studiengangpräferenz erhoben. Es wird außerdem sowohl in EG als auch in KG erfragt, welche sonstigen Informationsquellen in der Zwischenzeit genutzt wurden (ggf. auch OSA anderer Hochschulen, hochschulübergreifende OSA u. ä.). Die allgemeine Hypothese unter Annahme der Wirksamkeit des OSA lautet, dass die EG zum Posttestzeitpunkt ein höheres Vorwissen über und eine höhere Eignung und Passung zu den präferierten Studiengängen aufweist im Vergleich zur KG und im Vergleich zum Prätest-Zeitpunkt. Außerdem wird erwartet, dass diejenigen, die ihre Präferenz gewechselt haben, eine schlechtere Passung und Eignung zu dem ursprünglich präferierten Studiengang aufweisen.

Die theoretische Betrachtung und empirische Überprüfung möglicher Konfundierungen im Rahmen der summativen Evaluation kann außerdem einen Mehrwert für die Optimierung von OSA haben. Erkenntnisse darüber, welche Aspekte der individuellen Ausgangslage einen Einfluss auf die OSA-Nutzung und deren Wirksamkeit haben, können dann gezielt bei der Gestaltung von OSA berücksichtigt werden. So können Erkenntnisse über potentielle Konfundierungen eingesetzt werden, um die Nutzung und Wirkung von OSA langfristig zu verbessern. Die s Box „Beispielfragen zum Einfluss der Ausgangslage auf die Nutzung und Wirkung von OSA“ zeigt, welche Fragen ggf. mit Hilfe systematischer Untersuchungen zum Einfluss der Ausgangslage auf die Nutzung und Wirksamkeit von OSA beantwortet werden können.

### **Beispielfragen zum Einfluss der Ausgangslage auf die Nutzung und Wirkung von OSA**

- Welche Informationen braucht die Zielgruppe, um sich für die Teilnahme an einem OSA zu entscheiden?
- An welchen Stellen platziert man wirksam Werbung für das OSA? Welche Wege können wie genutzt werden, um die Zielgruppe adäquat anzusprechen und zur Teilnahme am OSA zu bewegen?
- Wie kann man bei der Gestaltung der Elemente eine nicht ernsthafte Bearbeitung durch wenig gewissenhafte Personen vermeiden (z. B. durch den Aufforderungscharakter der Instruktion)?
- Wie kann man Personen zur verbesserten Reflektion anleiten?
- Wie kann man das OSA gestalten, um besser mit bestehenden Vorannahmen und bereits vorhandenen Prioritäten der Zielgruppe umzugehen und diese in die Empfehlungen einzubinden?

---

#### 9.4.2 Bestimmung einer geeigneten Grundgesamtheit

Unter dem Begriff der Grundgesamtheit (auch: Population) versteht man die Menge aller potentiellen Personen, aus denen zur Durchführung einer Studie die Stichprobe gezogen werden soll. Im Rahmen der Untersuchungsplanung ist die Frage einer geeigneten Grundgesamtheit ebenfalls a priori zu klären. Für die Festlegung auf eine geeignete Grundgesamtheit, aus der Stichproben zur empirischen Überprüfung der Wirksamkeit von OSA gezogen werden können, muss die OSA-Zielgruppe unbedingt mitberücksichtigt werden. Zielgruppe eines typischen OSA für einen einzelnen Studiengang sind die Studieninteressierten für diesen Studiengang. Jedoch nehmen zum einen nicht alle Studieninteressierten an dem OSA teil. Für die Evaluation der Frage, ob das OSA die Zielgruppe erreicht, müssen so auch diejenigen berücksichtigt werden, die nicht am OSA teilgenommen haben. Zum anderen beginnen nicht alle OSA-Teilnehmerinnen und -Teilnehmer das jeweilige Studium. Dennoch basieren viele Studien zur Evaluation von OSA anhand klassischer Studienerfolgskriterien auf der Population der Studierenden an der spezifischen Hochschule und in dem spezifischen Studiengang. Ein OSA zielt aber mit seinen Wirkmechanismen auch darauf ab, die Studienwahlentscheidung aller Studieninteressierten, d. h. aller OSA-Teilnehmerinnen und Teilnehmer, zu verbessern (s. erfolgsvermittelnde Faktoren I in Abb. 9.1). In diesem Sinne kann sich die Wirksamkeit eines OSA auch in einer Entscheidung gegen einen bestimmten Studiengang oder für einen anderen Bildungsweg zeigen, wenn das OSA eine entsprechende fehlende Passung oder Eignung nahelegt. Die OSA-Zielgruppe und die gewählte Grundgesamtheit für empirische Evaluationsstudien sollten daher möglichst deckungsgleich sein, was in Abb. 9.2 auch graphisch veranschaulicht wird. Erfasst man zur Wirksamkeitsevaluation von OSA umgekehrt lediglich die späteren Studierenden in dem spezifischen Studiengang, kann das Ergebnis in mehrerlei Hinsicht verfälscht sein. Eine Stichprobe aus der Grundgesamtheit der Studierenden ist möglicherweise homogener im Hinblick auf deren individuelle Ausgangslage vor dem Studium als Studieninteressierte, die das OSA absolviert haben und das Studium aber nicht aufnehmen oder frühzeitig abbrechen. s Kap. 4 dieses Buches geht ebenfalls vor dem Hintergrund der Aussagekraft von Validierungsstudien für die im OSA genutzten Selbsttests auf diese Problematik ein.

#### 9.4.3 Wahl einer geeigneten Operationalisierung zur Vermeidung von Asymmetrie

Im Rahmen der Untersuchungsplanung wird auch festgelegt, welche Instrumente zur Erfassung der Personenmerkmale der Ausgangslage und der verschiedenen Kriterien (sowohl der erfolgsvermittelnden Faktoren als auch der Studienerfolgskriterien) eingesetzt werden, d. h. welche Indikatoren zur Erfassung genutzt werden. Betrachtet man die relevanten Personenmerkmale im Kontext von OSA (z. B. kognitive Fähigkeiten, Gewissenhaftigkeit, Schulnoten etc.) oder auch die Evaluationskriterien, so sind diese häufig hierarchisch konzipiert und können sich im Hinblick auf ihr Generalitätsniveau unterscheiden. So kann die generellere Persönlichkeitsdimension Gewissenhaftigkeit nach Soto & John (2017) in die spezifischeren Subfacetten Ordnungsliebe, Fleiß und Verlässlichkeit unterteilt werden. Bei den Evaluationskriterien setzt sich die übergeordnete Dimension des Studienerfolgs aus den spezifischeren Komponenten wie Noten, erworbenen ECTS oder der Intention zum Studienabbruch zusammen. Nicht zuletzt können sich die im OSA adressierten Wirkmechanismen und die sich daraus ergebende Wirkung im Hinblick auf das Generalitätsniveau unterscheiden. So adressieren studiengangübergreifende OSA ggf. andere Wirkmechanismen und globalere Effekte als studiengangspezifische OSA mit einer spezifischeren Passungs- und Eignungsbeurteilung.

Bei der Wahl geeigneter Indikatoren ist neben einer ausreichenden diagnostischen Güte (s. Kap. 5) auch darauf zu achten, dass die Messebenen der verschiedenen Instrumente den theoretischen Überlegungen nach symmetrisch sind, d. h. die passenden hierarchischen Ebenen adressieren. Das Prinzip der Symmetrie geht auf den theoretischen Ansatz von Brunswik (1955) zurück und „muss auf die paarweise Beziehung von Ursachen und Effekten [...] angewendet werden“ (Wittmann, 1990, S. 247). Bezogen auf den Kontext OSA kommen als Ursachen sowohl die Bearbeitung des OSA als auch die Personenmerkmale der Ausgangslage in Frage, unter die

---

Effekte fallen alle sowohl erfolgsvermittelnde Faktoren als auch Studienerfolgskriterien von OSA. Abb. 9.2 zeigt die drei Variablenbereiche und die wechselseitigen Beziehungen zwischen Ursachen und Effekten, bei denen Asymmetrie eine Bedrohung für die Interpretierbarkeit von empirischen Befunden darstellt. Typische Gefahren der Asymmetrie bestehen zum einen zwischen der Ausgangslage und den Kriterien. Beispielsweise sagt die Schulnote im Fach Mathematik als Indikator für das Personenmerkmal Vorwissen die Leistung in einer Statistiklausur besser vorher als die generellere Abiturnote dazu in der Lage ist. Zum anderen kann Asymmetrie auch zwischen den angenommenen Wirkmechanismen und den Evaluationskriterien auftreten. Darunter versteht man grundsätzlich, dass die Wirkung eines OSA nicht genau die angelegten Evaluationskriterien adressiert, sondern breiter oder enger ausgelegt sein kann. Ein allgemeines Studienorientierungs-OSA, welches aufgrund von Interessensprofilen Empfehlungen für mögliche Studiengänge ausspricht, wirkt so möglicherweise nicht spezifisch genug auf die Passung zu und Eignung für einen bestimmten Studiengang.

Liegt eine Asymmetrie zwischen Ursachen und Effekten vor, wie sie in den obigen Beispielen beschrieben ist, kann es zu Fehlinterpretation von empirischen Befunden zum Zusammenhang von Ursachen und Effekten kommen. Es wird angenommen, dass dieser Zusammenhang nur dann maximal ist, wenn die Indikatoren im Hinblick auf das Generalitätsniveau symmetrisch sind. Umgekehrt bedeutet dies, dass der Zusammenhang umso mehr unterschätzt wird, je mehr eine Asymmetrie gegeben ist.

## 9.5 Fazit und Ausblick

Die Evaluation von OSA sollte ein unabhängiger Prozess mit klaren Verantwortlichkeiten sein, der von Anfang einer OSA-Entwicklung an mitgedacht werden sollte. Dabei ist es für die Aussagekraft und Nützlichkeit der Evaluationsergebnisse von zentraler Bedeutung, dass die methodischen Anforderungen des jeweiligen Evaluationsansatzes Berücksichtigung finden (v. a. die Evaluations- bzw. Qualitätsstandards). Für den Nutzen einer Evaluation ist dabei besonders ein kriteriengeleitetes Vorgehen wichtig, das sich an den am OSA Beteiligten orientiert. Um die Nutzbarkeit und Nützlichkeit der Ergebnisse zu gewährleisten, ist außerdem wichtig, dass potentielle Limitationen der Aussagekraft von Studienergebnissen im Rahmen der Berichtstellung angemessene berücksichtigt werden.

Bisherige empirische Studien zur Evaluation von OSA befassen sich selten mit Fragen der summativen Evaluation. Die Frage, ob und in welchem Ausmaß ein OSA tatsächlich die gewünschte Wirkung erzielt, ist jedoch zentral für eine langfristige Implementierung von OSA in der deutschsprachigen Hochschullandschaft. Gerade bei der summativen Evaluation von OSA lassen sich methodische Anforderungen identifizieren, die bislang nur wenig Berücksichtigung fanden und die gleichzeitig jedoch zentral für die Aussagekraft solcher Studien sind: Auf Basis theoretischer Überlegungen sollten solche Evaluationskriterien gewählt werden, die im direkten Zusammenhang mit den angenommenen Wirkmechanismen des OSA stehen, so dass eine faire Überprüfung der Wirksamkeit gewährleistet werden kann. Für die Wahl eines adäquaten Forschungsdesigns sollten ebenfalls auf Basis theoretischer Überlegungen Störvariablen identifiziert werden, die mit den Wirkmechanismen des OSA und den jeweiligen Kriterien konfundiert sind. Um potentielle Konfundierung zu vermeiden oder zu kontrollieren, sollten die notwendigen untersuchungsplanerischen Maßnahmen ergriffen werden (Kontrollgruppe, Randomisierung oder Parallelisierung, Prä-Post Forschungsdesign). Ebenfalls im Rahmen der Untersuchungsplanung sollte wohlüberlegt sein, welche Grundgesamtheit für die Studie vor dem Hintergrund der OSA-Zielgruppe relevant ist (Studieninteressierte, OSA-Teilnehmerinnen und – Teilnehmer, Studierende o. a.), um die Wirkung von OSA auch an der anvisierten Zielgruppe zu untersuchen. Zudem sollten die Indikatoren zur Erfassung von Merkmalen der Ausgangslage, Störvariablen und Kriterien so ausgewählt werden, dass Symmetrie zwischen Ursachen- und Effektvariablen gegeben ist.

Betrachtet man die im vorliegenden Kapitel aufgezeigten Anforderungen an die Planung von Evaluationsstudien, so scheint die Evaluationsforschung zu Online Self-Assessments noch in den Kinderschuhen zu stecken. Dabei kann man nicht genug den Mehrwert von methodisch

---

durchdachten Evaluationsstudien betonen. Unabhängig von den erzielten Ergebnissen stärkt eine umfassende Evaluation nicht nur das jeweilige OSA, sondern bietet auch Ansatzpunkte für weiterführende Forschungsfragen (s. Box „Beispielfragen zum Einfluss der Ausgangslage auf die Nutzung und Wirkung von OSA“). Die Beantwortung solcher Fragen kann nicht nur zu einem besseren Verständnis des Prozesses der Studienfachwahl oder der daraus resultierenden Konsequenzen für ein erfolgreiches Studium beitragen. Sie bietet auch die Möglichkeit, Stellschrauben für eine Optimierung der Intervention Online Self-Assessment zu identifizieren.

### Checkliste

1. Wurde die Evaluation des OSA rechtzeitig in die Planungen einbezogen?
2. Wurden alle im Laufe des Evaluationsprozesses zu klärenden Fragen – unter Berücksichtigung der methodischen Anforderungen – geklärt?
  - a. Was sind die Rahmenbedingungen der Evaluation? (Zweck und Rolle der Evaluation, Interessensgruppen, zeitlicher Rahmen, Evaluationsgegenstand, Machbarkeit)
  - b. Welcher Evaluationsansatz wird verfolgt? (Summative vs. formative Evaluation)?
  - c. Auf welchem Wege sollen die Evaluationskriterien bestimmt werden? Wird bei der Bestimmung der Kriterien transparent und beteiligtenorientiert vorgegangen?
  - d. Werden bei der Planung, Durchführung und Auswertung der Evaluationsstudie die methodischen Anforderungen berücksichtigt?
  - e. Werden mögliche Limitationen für die Aussagekraft der Evaluationsstudie in adäquater Form den Nutzerinnen und Nutzer der Evaluationsergebnisse vermittelt?
3. Wurde bei der Auswahl der Kriterien zur summativen Evaluation deren Zusammenhang mit den angenommenen Wirkmechanismen des OSA berücksichtigt?
4. Wurden ausreichende Maßnahmen zur Sicherstellung der internen Validität des Forschungsdesigns ergriffen? (Vermeidung oder Kontrolle von Konfundierung)
5. Wurde bei der Wahl der Grundgesamtheit für die summative Evaluationsstudie die OSA-Zielgruppe berücksichtigt?
6. Wurde bei der Wahl von Indikatoren zur Erfassung von Merkmalen der Ausgangslage, Störvariablen und Kriterien auf die Symmetrie zwischen Ursachen- und Effektivariablen geachtet?
7. Ergeben sich aus den vorliegenden Evaluationsergebnissen weiterführende Forschungsfragen?

## 9.6 Literatur

- Balzer, L. (2005). *Wie werden Evaluationsprojekte erfolgreich? – Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess*. Landau: Verlag Empirische Pädagogik.
- Beauducel, A. & Leue, A. (2014). *Psychologische Diagnostik*. Göttingen: Hogrefe-Verlag.
- Bildung & Begabung (2021). *Review Online-Self-Assessments an staatlichen Hochschulen in Deutschland*. Zugriff am 14.3.2021. Verfügbar unter: <https://www.begabungslotse.de/specials/check-wunschstudium/studiendesign>
- Beywl, W. & Widmer, T. (2009). *Evaluation in Expansion: Ausgangslage für den intersektoralen Dreiländer-Vergleich*. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation* (S. 13–23). Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-91468-8>
- Brunner, S. (2017). *Online-Self-Assessments*. Carl von Ossietzky Universität Oldenburg: Koordinierungsstelle der Begleitforschung des Qualitätspaktes Lehre (KoBF). Wolfgang Schulenberg-Institut für Bildungsforschung und Erwachsenenbildung.
- Brunner, S., Kretschmer, S. & Maschwitz, A. (2017). *Online-Self-Assessments an deutschen Hochschulen*. In K. Mayrberger (Hrsg.), *Universitätskolleg-Schriften* (Band 23, S. 181–234). Universität Hamburg. Verfügbar unter: <https://www.universitaetskolleg.uni-hamburg.de/publikationen/uk-schriften-023.pdf>
- Brunswick, E. (1955). *Representative design and probabilistic theory in a functional psychology*. *Psychological Review*, 62(3), 193–217. <https://doi.org/10.1037/h0047470>
- DeGEval – Gesellschaft für Evaluation e.V. (2017). *Standards für Evaluation*. Erste Revision 2016. Mainz: DeGEval. Verfügbar unter: <https://www.degeval.org/degeval-standards/standards-fuer-evaluation/>

- 
- DIN Deutsches Institut für Normierung e.V. (2016). DIN 33430. Anforderungen an berufsbezogene Eignungsdiagnostik. Berlin: Beuth Verlag GmbH. <https://doi.org/10.31030/2514220>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. (5. Auflage.). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>
- Eid, M., Gollwitzer, M. & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (4. Auflage.). Weinheim: Beltz.
- Gollub, J. & Meyer-Guckel, D. V. (2014). Wer bin ich – und wenn ja, wie viele? Online-Studienselbsttests als „Orientierungs- und Entscheidungshelfer“. *Wissenschaftsmanagement*, 5, 36–41.
- Hattie, J. A. C. (2013). *Lernen sichtbar machen*. Überarbeitete deutschsprachige Ausgabe von „Visible learning“, besorgt von W. Beywl & K. Zierer, Baltmannsweiler: Schneider Verlag Hohengehren.
- Heukamp, V., Putz, D., Milbradt, A. & Hornke, L. F. (2009). Internetbasierte Self-Assessments zur Unterstützung der Studienentscheidung. *Zeitschrift für Beratung und Studium. Handlungsfelder, Praxisbeispiele und Lösungskonzepte.*, 4(1), 2–8.
- Ibabe, I. & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59(2), 243–258. <https://doi.org/10.1007/s10734-009-9245-6>
- Kubinger, K. D., Frebort, M. & Müller, C. (2012). Self-Assessment im Rahmen der Studienberatung: Möglichkeiten und Grenzen. In K.D. Kubinger, M. Frebort, L. Khorramdel & L. Weitensfelder (Hrsg.), *Self-Assessment: Theorie und Konzepte* (S. 9–24). Lengerich: Pabst Science Publ.
- Milbradt, A., Zettler, I., Putz, D., Heukamp, V. & Hornke, L. F. (2008). Ziele von Self-Assessments und ihre Bedeutung für Entwicklung und Evaluation. *Report Psychologie*, 33, 352–362.
- Ott, M., Ulfert, A. & Kersting, M. (2017). "Online-Assessments" und "Self-Assessments" in der Eignungsdiagnostik. In: D.E. Krause (Hrsg.); *Personalauswahl* (S. 215–242). Wiesbaden: Springer Fachmedien: DOI 10.1007/978-3-658-14567-5\_10
- Schmitt, M. & Schmidt-Atzert, L. (2019). Gütekriterien für Online-Self-Assessments. In L. Schmidt-Atzert, M. Schütz & G. Stemmler (Hrsg.), *Online-Self-Assessments an Hochschulen* (S. 99–114). Pabst.
- Soto, C. J. & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143. <https://doi.org/10.1037/pspp0000096>
- Stockmann, R. (2004). Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren. CEval-Arbeitspapier, 9. Saarbrücken: Universität des Saarlandes, Fak. 05 Empirische Humanwissenschaften, CEval - Centrum für Evaluation.
- Thiele L. & Kauffeld S. (2020) Online Self-Assessments zur Studien- und Universitätswahl. In: S. Kauffeld & D. Spurk (Hrsg.) *Handbuch Karriere und Laufbahnmanagement*. Springer Reference Psychologie. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-662-45855-6\\_4-2](https://doi.org/10.1007/978-3-662-45855-6_4-2)
- Trapmann, S., Hell, B., Hirn, J.-O. W. & Schuler, H. (2007). Meta-Analysis of the Relationship Between the Big Five and Academic Success at University. *Zeitschrift für Psychologie*, 215(2), 132–151. <https://doi.org/10.1027/0044-3409.215.2.132>
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21(1), 11–27. <https://doi.org/10.1024/1010-0652.21.1.11>
- Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der Fünf-Datenboxen—Ein Rahmenkonzept für umfassende Evaluationsforschung. *Zeitschrift für Pädagogische Psychologie*; 4(4), 241–251.
- Zimmerhofer, A., Heukamp, V. M. & Hornke, L. F. (2006). Ein Schritt zur fundierteren Studienfachwahl - webbasierte Self-Assessments in der Praxis. *Report Psychologie*, 31(2), 62–72.