

Validität biographischer Daten zur Vorhersage von Trainings- und Berufserfolg: Ergebnisse einer Metaanalyse

Dipl.-Psych. Steffen Reinink

Um die Eignung eines Kandidaten für eine Tätigkeit abschätzen zu können, bieten sich Personalverantwortlichen und Wissenschaftlern drei methodische Zugänge an, die situationsübergreifend hohe kriteriumsbezogene Validität aufweisen: (a) der Eigenschaftsansatz, (b) der Simulationsansatz und (c) der biographische Ansatz (Schuler & Höft, 2006). Die vorliegende Metaanalyse konzentriert sich auf den letztgenannten Ansatz und stellt dabei einen Beitrag zur Klärung der Frage dar, wie verlässlich und ertragreich biographische Daten in unterschiedlichen methodischen und situativen Kontexten sind. Konkret besteht das primäre Ziel dieser Arbeit in der artefaktbereinigten Schätzung der Validität biographischer Daten zur Vorhersage von Trainings- und Berufserfolg. Derartige Validitätsschätzungen sind im eignungsdiagnostischen Kontext von höherem Interesse als theoretische Validitäten unter der Annahme perfekter Messgenauigkeit. Gegenwärtig mangelt es jedoch an Studien, welche die operationale Validität biographischer Daten dokumentieren. Zwar berichten Schmidt und Hunter (1998) operationale Validitäten für Berufserfolg, differenzieren dabei aber nicht zwischen objektiven und subjektiven Erfolgsindikatoren. Hinzu kommt, dass ihre Angaben für das Kriterium Trainingserfolg noch auf Ergebnissen aus dem Jahr 1984 basieren. Systematischer und umfangreicher untersuchte Bliesener im Jahr 1996 die Validität biographischer Daten für diverse Kriterien. Jedoch nahm er dabei keine Artefaktkorrekturen vor. Das Ziel der vorliegenden Arbeit besteht deshalb darin, diese Lücke in der Arbeit Blieseners zu schließen und zur Vergleichbarkeit mit anderen eignungsdiagnostischen Ansätzen operationale Validitätskoeffizienten zu schätzen. Ein weiteres Ziel dieser Arbeit ist, die zentralen methodischen sowie situativen Moderatoren der Validität biographischer Daten zu identifizieren. Zusätzlich zu den acht von Bliesener berücksichtigten Einflussgrößen (a) Untersuchungsdesign, (b) Validierungstechnik, (c) Konstruktintegrität eines Prädiktors, (d) Spezifität eines Prädiktors, (e) Zufallssignifikanzen (*capitalizing on chance*), (f) Berufsgruppe, (g) Kategorie des Kriteriums (Training vs. Beruf) und (h) Geschlecht sollen weitere potentielle Moderatoren untersucht werden, nämlich (i) die Art der Operationalisierung des Kriteriums (objektiv vs. subjektiv), (j) die Konstruktionsart des Prädiktors, (k) die Branche und (l) die Komplexität einer Aufgabe.

Methode

Als Definition biographischer Daten lag dieser Arbeit die Publikation Maels aus dem Jahr 1991 zu Grunde. Auf der Basis früherer Taxonomien benannte Mael zehn bipolare Eigenschaften, anhand derer biographische Items klassifiziert werden können. Um sicherzustellen, dass alle Arbeiten, welche in die vorliegende Metaanalyse integriert wurden, eine vergleichsweise hohe Konstruktvalidität aufwiesen, wurde jedes Primärstudien-Item entsprechend der Taxonomie Maels (1991) hinsichtlich seiner biographischen Qualität beurteilt. Als biographische Prädiktoren galten Items dann, wenn sie sieben von zehn Eigenschaften der oben genannten Taxonomie aufwiesen. Als einziges zwingend notwendiges Merkmal für biographische Daten galt allerdings das der Historizität. Nur Prädiktoren, die mehrheitlich aus biographischen Items bestanden, fanden in dieser Arbeit Berücksichtigung.

Eine umfangreiche Literaturrecherche in mehreren Datenbanken und Bibliographien lieferte mehr als 420 Treffer, aus denen letztlich 62 internationale Primärstudien als Datenbasis resultierten. Aus diesen Studien konnten insgesamt 115 Untersuchungseinheiten isoliert werden, die statistisch voneinander unabhängige Teiluntersuchungen darstellten. Daran anschließend wurden nach Sichtung mehrerer aktueller Metaanalysen ein Kodierschema sowie ein Kodiermanual entwickelt. Das *Kodiermanual* enthielt Anweisungen und Regeln über die Kodierung und diente als Maßgabe und Entscheidungsgrundlage im gesamten Prozess der Datenextraktion. Das *Kodierschema* diente der Strukturierung und Sammlung der Informationen aus den Primärstudien. Zur Schätzung der Reliabilität des Kodiervorgangs wurden mit Hilfe einer unabhängigen Zweitkodierung auf metrischem Skalenniveau Intraklassen-Korrelationskoeffizienten und auf nominalem Niveau Cohens Kappas berechnet. Die Schätzung ergab für die Art und Höhe der Validitätskoeffizienten, für die Stichprobengröße, das Geschlecht, die Berufshauptgruppe, den Fachkenntnisgrad (*skill level*), die Konstruktintegrität sowie für die Variable Zufallssignifikanz perfekte Übereinstimmungen (1.0). Substantielle Übereinstimmungen lagen für die Kodierung der Branche (.735) und der Prädiktorspezifität (.640) vor. Moderat fiel die Interrater-Reliabilität hinsichtlich des Untersuchungsdesigns aus (.400). Nur gering war die Übereinstimmung für die Kodierung der Prädiktorvalidierungstechnik (.122). Hinsichtlich der Prädiktor- und Kriteriumsreliabilitäten sowie der Varianzeinschränkungskoeffizienten konnte mangels Primärstudieninformationen keine Übereinstimmung bestimmt werden.

Die Integration der Primärstudienbefunde geschah im Anschluss an die Kodierung in Form einer Metaanalyse über Zusammenhangsmaße unter Gebrauch von Artefaktverteilungen. Diese metaanalytische Variante wurde gerechnet, weil es in den einzelnen Primärstudien oftmals

an Informationen über psychometrische Gütekriterien sowie Artefakte mangelte. Um für die Berechnung das adäquate metaanalytische Modell zu bestimmen, wurde kein Homogenitätstest durchgeführt. Vielmehr wurde a priori ein Modell zufälliger Effekte angenommen. Diese Entscheidung basierte auf den Empfehlungen des *National Research Councils (1992)*, Modelle fester Effekte generell durch Modelle zufälliger Effekte zu ersetzen. Jeder beobachtete Validitätskoeffizient wurde zur Schätzung der operationalen Validitäten im Einklang mit anderen aktuellen Metaanalysen um drei Artefakte korrigiert: (a) den Stichprobenfehler, (b) die direkte Varianzeinschränkung und (c) Messfehler in Kriterium und Prädiktor. Letztgenannte Korrektur diente jedoch (einzig) einer verlässlichen und vergleichbaren Schätzung des durch Artefakte erklärten Varianzanteils. Abschließend erfolgte die Testung der Moderatoreffekte gemäß den Empfehlungen von Hunter und Schmidt (2004) für a priori Fragestellungen. Diese Empfehlung sieht vor, die einzelnen Moderatorstufen in Form von Subgruppen zu kodieren, separat auszuwerten und darauf folgend Konfidenzintervalle zu bestimmen.

Ergebnisse

Sowohl für Berufserfolg als auch für das Kriterium Trainingserfolg lag die Anzahl der Untersuchungseinheiten mit $k = 60$ für Beruf und $k = 59$ für Training recht hoch. Deutlich weniger ($k = 39$) Untersuchungseinheiten flossen in die Schätzung der Validität biographischer Daten zur Vorhersage von Noten ein. Gleichermaßen variierte die totale Stichprobengröße über die drei Kriterien hinweg mit einem Maximum von $N = 48744$ für Training und einem Minimum von $N = 20088$ für Noten. Die totale Stichprobengröße für das Kriterium Berufserfolg lag bei $N = 43588$. Da die berechneten Konfidenzintervalle für keines der Kriterien den Wert 0 mit einschlossen, bestand für jedes Kriterium ein statistisch bedeutsamer Zusammenhang mit biographischen Daten. Überdies überschritten sich die 95 %-Konfidenzintervalle nicht. Daraus kann gefolgert werden, dass sich die Validität biographischer Daten in statistisch signifikanter Weise für die drei Kriterien unterscheidet. Mit einer korrigierten Validität von $\rho = .279$ lässt sich Berufserfolg mit Hilfe biographischer Daten schlechter vorhersagen als Erfolgsindikatoren, die sich auf den Ausbildungskontext beziehen (Trainingserfolg: $\rho = .420$; Noten: $\rho = .595$).

Anhand der Analyse der Subgruppen konnten kriterienübergreifend und in Übereinstimmung mit Bliesener (1996) Moderatoreffekte für das Untersuchungsdesign, die Validierungstechnik und die Spezifität eines Prädiktors festgestellt werden. Im Gegensatz zu Bliesener (1996) konnte für die Variablen Konstruktintegrität und Zufallssignifikanz (*capitalizing on chance*) keine moderierende Wirkung gefunden werden. Die Annahmen Blieseners (1996) über die moderierende Wirkung des Geschlechts und des Berufs konnten in der vorliegenden Arbeit

nur in Teilen bestätigt werden. Für beide potentiellen Moderatoren ergaben sich uneinheitliche Ergebnisse in Bezug auf die unterschiedenen Kriterien. Für Berufserfolg, nicht aber für Trainingserfolg wurde ein Moderatoreffekt der Berufshauptgruppe gefunden. Die Geschlechter wiesen im Gegensatz dazu nur für Trainingserfolg und nicht für Berufserfolg differentielle Validitäten in statistisch bedeutsamer Weise auf. Ein eindeutiges Resultat ergab indes die Untersuchung über die moderierende Wirkung der Branche, denn die Validität biographischer Daten kann sowohl für Trainingserfolg auch als für Berufserfolg über Wirtschaftssektoren hinweg generalisiert werden. Was den Komplexitätsgrad einer Aufgabe betrifft, so konnte nur für das Kriterium Berufserfolg ein Moderatoreffekt gefunden werden. Entgegengesetzt hierzu wiesen biographische Daten nur im Ausbildungskontext verschiedene Validitäten für subjektiv versus objektiv gemessenen Erfolg auf.

Diskussion

Generell belegen die Befunde die Validität biographischer Daten zur Vorhersage von Erfolg. Im Vergleich mit den Ergebnissen anderer Forscher fielen die Validitätsschätzungen dieser Arbeit jedoch etwas niedriger aus. Diese Differenzen können wahrscheinlich auf methodische Faktoren zurück geführt werden. So basierte zum einen die vorliegende Arbeit auf einer größeren Datenmenge als andere Metaanalysen über die Validität biographischer Daten. Zum anderen wurde in dieser Arbeit ein vergleichsweise konservatives Vorgehen umgesetzt. Beispielsweise musste gelegentlich zur Kombination von Koeffizienten innerhalb der Primärstudien auf Fishers Z-Werte zurückgegriffen werden, weil Angaben über die Interkorrelationen der Prädiktor- und Kriteriumskomponenten fehlten. Ebenso resultierte die Aufnahme einiger kategorialer Zusammenhangsmaße zur Erhöhung der Teststärke in einer konservativeren Validitätsschätzung. Ein weiteres Beispiel für das konservative Vorgehen betrifft die Moderatoranalysen. Um den Fehler erster Ordnung bei der Untersuchung der Moderatoren möglichst gering zu halten, wurden t-Verteilungen anstelle der Standardnormalverteilung bei der Berechnung der Konfidenzintervalle genutzt. Zusammenfassend kann hieraus geschlossen werden, dass die Befunde dieser Arbeit die wahre Validität biographischer Daten tendenziell eher unterschätzen.

Betrachtet man die Ergebnisse im Gesamtzusammenhang beruflicher Eignungsdiagnostik, lässt sich sagen, dass die Validität biographischer Daten als hoch zu beurteilen ist. Hinsichtlich der Vorhersage von Trainingserfolg stellen biographische Daten direkt nach Verfahren zur Messung allgemeiner kognitiver Fähigkeiten den stärkste Prädiktor dar (vgl. hierzu Robertson & Smith, 2001, S. 442). Was die Vorhersage von Berufserfolg betrifft weisen allerdings sowohl Konstruktorientierte Verfahren als auch Instrumente des

Simulationsorientierten Ansatzes höhere Validitäten auf als biographische Prädiktoren. Dass Biographieorientierte Verfahren dennoch auch in der beruflichen Eignungsdiagnostik regelmäßig eingesetzt werden sollten, liegt insbesondere an ihrer hohen inkrementellen Validität. Gerade in der Kombination mit anderen Instrumenten entfalten biographische Daten ihre Wirkung und verbessern die Vorhersage von Berufserfolg deutlich (Stokes & Cooper, 2004). Für die Praxis implizieren die Befunde dieser Arbeit jedoch, dass biographische Daten nicht unreflektiert eingesetzt werden sollten, weil ihre kriteriumsbezogene Validität durch oben genannte Moderatoren beeinflusst wird. Aus diesem Grund sollte im Vorfeld stets geprüft werden, ob Biographieorientierte Verfahren im jeweiligen situativen Anwendungskontext valide Erfolgsprädiktoren sind.

Literaturverzeichnis

- Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology*, 69, 107–120.
- Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in reserach findings*. Newbury Park, CA: Sage.
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44, 763–792.
- National Research Council (1992). *Combining information: Statistical issues and opportunities for research*. Washington: National Academy Press.
- Robertson, I. T. & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, 74, 441-472.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schuler, H. & Höft, S. (2006). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 101-144). Göttingen: Hogrefe.
- Stokes, G. S. & Cooper, L. A. (2004). Biodata. In J. C. Thomas & M. Hensen (Hrsg.), *Comprehensive handbook of psychological assessment* (Bd. 4: Industrial and organizational assessment, S. 243-268). Hoboken, NJ: Wiley.