

# Explikation von Entscheidungsregeln in der Personalauswahl

Dipl.-Psych. Julia Edler

Die Hauptaufgabe der Experten in der Eignungsdiagnostik ist es, aus den gesammelten Informationen über den Bewerber zu einem Urteil zu kommen, das die Eignung bzw. Passung des Bewerbers für eine spezielle Stelle ausdrückt. Wie aber integriert der Experte dieses Wissen zu einem Gesamturteil?

Studien über das Wissen von Experten zeigen, dass Expertenwissen häufig in impliziter Form vorliegt (vgl. Büssing, Herbig & Latzel, 2004). Experten sind zwar in der Lage, eine Eignungsbeurteilung durchzuführen, können ihr Wissen und Vorgehen aber selten beschreiben; sie handeln erfahrungsgelenkt (Rose, 1992). Damit entziehen sich ihre Auswahlstrategien einer Prüfung der Gütekriterien, Gleichstellungsansprüchen und eventuell anderer problematischer Inhalte. Für eine wissenschaftliche und den Qualitätsstandards der DIN 33430 (NAGD, 2002) entsprechenden Eignungsdiagnostik sollten aber explizite, transparente Entscheidungsstrukturen vorliegen (Westhoff, Hellfritsch, Hornke, et al., 2006). Nur so können Auswahlstrategien effizient angewandt, vermittelt und weiterentwickelt werden. Außerdem ist die empirische Festlegung von Entscheidungsregeln in Personalauswahlverfahren auch eine Legitimation gegenüber den betrieblichen Entscheidungsträgern (vgl. Höft, 2006, S. 762) und sollte deshalb durch abgesicherte, diskriminierungsfreie Regeln bestätigt werden.

Die vorliegende Arbeit beschäftigt sich mit der Formalisierung dieser Entscheidungsregeln am Beispiel einer Auswahlsequenz für ein Traineeprogramm und untersucht mögliche Methoden zur Gewinnung empirischer Entscheidungsregeln. Für die Ableitung dieser impliziten Regeln stehen zwei Paradigmen zur Verfügung: (1) der strukturelle multivariat-statistische Ansatz (Brehmer, 1994) und (2) der Process-Tracing-Ansatz (Harte & Koele, 1995). Beim multivariat-statistischen Ansatz wird die Beziehung zwischen der gegebenen Information (dem Prädiktor) und der Entscheidung (dem Kriterium) durch die Anwendung eines optimal angepassten Algorithmus expliziert. Beim Process-Tracing-Ansatz liefern die kognitiven Prozesse der Experten die Basis zur Extraktion der Entscheidungsregeln. Das typische Vorgehen dabei ist, den Experten aufzufordern, anhand vorgelegter Informationen ein Urteil zu treffen und dieses durch ein Modell nachzubilden. Prozessmodelle greifen dieses Vorgehen auf, indem sie versuchen, die wahre Entscheidungsstruktur in einem mathematischen Modell abzubilden.

Im Rahmen dieser Studie werden auf der Basis spezieller kognitiver Fähigkeitstest implizite Entscheidungsregeln von Personalauswahlexperten mit Hilfe des Process-Tracing-Ansatzes und des multivariat-statistischen Ansatzes hergeleitet und in ihrer Vorhersagegenauigkeit, Stabilität und Anwendbarkeit empirisch verglichen. Dabei soll die Hypothese getestet werden, ob Entscheidungsregeln durch explorative Prozessmodellierung im Rahmen des Process-Tracing-Ansatzes reliabler und valider als durch multivariate Methoden der statistischen Urteilsbildung expliziert werden können. Die dafür notwendigen Schritte teilen sich auf in eine Bewertung der Reliabilität bzw. Stichprobenunabhängigkeit, eine Überprüfung der Vorhersagegenauigkeit und eine direkte Gegenüberstellung der Modelle. Zusätzlich soll in

der Arbeit untersucht werden, ob sich durch Mehrheitsentscheidung bessere Klassifikationsergebnisse erzielen lassen.

## **Methode**

Für die Erstellung der Prozessmodelle wurden die für die Personalauswahl verantwortlichen Experten eines Unternehmens ( $N=7$ ) durch ein halb-strukturiertes Interview zu ihren Entscheidungsregeln befragt. Es wurden verschiedene Verfahren angewandt, um Varianten der Entscheidungsfindung abzudecken. Datengrundlage der statischen Analysen sind die Originaltestprofile ( $N=397$ ) der Bewerber.

**Lautes Denken.** Den Experten wurden nacheinander 29 den Originalprofilen nachempfundene Testprofile präsentiert. Sie wurden gebeten diese zu bewerten. Bei den Beurteilungen der Testprofile sollten sie ihre Gedanken verbalisieren, es wurde die Methode des „Lauten Denkens“ (vgl. Ranyard & Williamson, 2005) angewandt. Um die transkribierten Interviews einer quantitativen Analyse zugänglich zu machen, wurden diese zunächst auf der Basis eines Kategoriensystems kodiert und als Ausschlussregeln auf die Bewerberprofile angewandt.

**Fragebogen.** Alle Experten wurden gebeten einen Fragebogen auszufüllen. Der Fragebogen beinhaltet Fragen zur Entscheidungsstrategie (kompensatorisch vs. nicht-kompensatorisch) und Cut-off-Werten. Die unteren und oberen Cut-off-Werte wurden dann ohne weitere Kodierung auf alle Profile angewandt.

**Expertenregeln.** Im Anschluss an die Befragung wurden die Experten gebeten, ihre zuvor getroffenen Entscheidungen und Regeln zu reflektieren und durch Wenn-dann-Regeln zu erklären. Die von den Experten selbst erstellten Regeln wurden in Anlehnung an die Methode HYPAG (Wottawa, 1984) direkt in eine Datenmatrix eingegeben und die Ergebnisse den Experten zurückgemeldet. Aufbauend auf den noch falsch klassifizierten Fällen wurden die Experten aufgefordert, weitere Regeln zu formulieren.

**Erstellung der Gesamtmodelle.** Die drei Prozessmodelle: „Lautes Denken“, „Fragebogen“ und „Expertenregeln“ wurden für jeden der sieben Experten erstellt. Um drei Gesamtmodelle zu erzeugen, wurden Variablen erzeugt, die die Anzahl der positiven Entscheidungen pro Bewerber für jedes Modell zählen. War die Anzahl der übereinstimmenden Urteile größer oder gleich 4, wurde die Gesamtentscheidung positiv bzw. negativ ins Gesamtmodell übernommen. Es handelt sich also um eine klassische Mehrheitsentscheidung (vgl. Davis, 1973). Wenn sich mindestens vier Experten einig waren, steht die Entscheidung für das Gesamtmodell fest.

Im zweiten Schritt sollen die Bewerberprofile multivariat ohne Berücksichtigung des Expertenwissens ausgewertet werden. Es wurden folgende Modelle erstellt:

- Lineare Modelle - Diskriminanzanalyse
- Nichtlineare Modelle - Logistische Regression
- Konfigurale Modelle - Answer Tree/CHAID Analyse

Alle Modelle zeigen nach Anwendung der Regelsätze eine eindeutige Klassifizierung der Bewerberprofile und die Urteilsgenauigkeit an.

## Design zur vergleichenden Überprüfung der Explikationsmodelle

Der Vergleich der verschiedenen Modelle hinsichtlich der Vorhersagegenauigkeit konnte mittels einer Vierfeldertabelle und daraus abgeleiteter Kennzahlen vorgenommen werden. Diese Koeffizienten sind *Selektiver Eignungsquotient*, *Negativer Prädiktionswert*, *Fehler negativ*, *Fehler positiv*, *Sensitivität*, *Spezifität*, *Trefferquote*, *Fehlerquote*, *Phi-Koeffizient* und *Treffer-Fehler-Score* (angelehnt an Noack & Petermann, 1992, S.299).

In einem weiteren Schritt wurden die Modelle auf ihre Reliabilität und Stabilität überprüft. Durch einen Vergleich der ICC-Koeffizienten wurde ermittelt, ob die Prozessmodelle übereinstimmende Entscheidungsregeln hervorbringen. Die Prüfung auf Stichprobenunabhängigkeit der multivariat-statistischen Modelle wurde durch eine Kreuzvalidierung analysiert. So konnte für beide Modellarten (Prozess vs. multivariat-statistisch) das stabilste Modell ermittelt werden.

### Ergebnisse

#### Prozessmodelle

Auf der Basis der Einzelmodelle wurden für jeden Experten das Übereinstimmungsmaß Intra Class Correlation (ICC) mit Einweg-Zufallseffekten berechnet. Ziel der Berechnung war es, die Urteilsübereinstimmung der drei Explikationsmethoden festzustellen. Die Urteilskonkordanzen zeigen durchweg signifikante Übereinstimmungen. Insgesamt liegen die Konsistenzmaße der Urteile in 6 von 7 Fällen über dem von Bono und Judge (2003) angegebenen Richtwert von  $r=.30$ .

Wird die Gesamtgruppe der Experten in den Entscheidungsprozess einbezogen und aus den Einzelurteilen ein Gesamtmodell erstellt, ergeben sich insgesamt deutlich bessere Vorhersagen als das durchschnittliche Einzelmodell und häufig auch eine bessere als jedes Einzelmodell. Die Klassifikationsergebnisse der erstellten Gesamtmodelle zeigen, dass die Gütemaße ansteigen, wenn die Entscheidungsregeln unabhängig voneinander in die Gesamtentscheidung einfließen. Die positiven Eigenschaften unterschiedlicher Meinungen, vereint in einer Mehrheitsentscheidung, werden besonders beim Prozessmodell „Expertenregeln“ deutlich, welches als Einzelmodell eher schlecht in den Analysen abgeschnitten hatte. Durch die Anwendung einer einfachen Entscheidungsregel basierend auf der Theorie der sozialen Entscheidungsschemata (Davis, 1973) lässt sich die Effektivität der Bewerberklassifikation durch den Experten deutlich erhöhen. Durch die Mehrheitsentscheidung, in der die Entscheidung jedes einzelnen Experten einfließt, steigen die Gütemaße der Modelle an. Das Gesamtmodell stellt damit ein besseres Entscheidungsmodell für diese Auswahlsequenz dar, wobei die verschiedenen Prozessmodelle unterschiedliche Stärken zeigen.

Die Stärken dieses Ansatzes liegen in einer relativ genauen Abbildung der Originaltesturteile bei einem geringen Risiko, ungeeignete Bewerber fälschlicherweise anzunehmen und geeignete Bewerber abzulehnen. Die Stärken des Gesamtmodells „Fragebogen“ liegen in der hohen Wahrscheinlichkeit, einen geeigneten Bewerber als solchen zu identifizieren. Allerdings sollte dieser Parameter nicht ohne Sensitivität interpretiert werden, die in diesem Fall vergleichsweise niedrig ausfällt. Ebenfalls niedrig fallen Trefferquote, Phi-Koeffizient und Treffer-Fehler-Score aus. Damit erweisen sich die Gütemerk-

male des Prozessmodells „Fragebogen“ im Vergleich zu den anderen Prozessmodellen als nicht zufriedenstellend. Das Prozessmodell „Expertenregeln“ schneidet trotz einer vergleichsweise guten Trefferquote ebenfalls nicht zufriedenstellend ab, da es die niedrigste Spezifität und den niedrigsten Treffer-Fehler-Score zeigt. Insgesamt zeigen sich also durch die verschiedenen Explikationsmethoden qualitative Unterschiede in den Beurteilungen der Testprofile.

### **Multivariat statistische Modelle**

Eine Analyse der Kreuzvalidierung zeigt hohe Übereinstimmung in der Klassifikation der Bewerber. Die Ergebnisse weisen damit auf eine stabile Modellkonstruktion hin.

Die Stärken der Diskriminanzanalyse liegen in einer relativ genauen Abbildung der Originaltesturteile. Dieser Befund wird durch den sehr niedrigen Fehlerquotient positiv bestätigt. Die Stärken der logistischen Regression liegen in der hohen Trefferquote, der niedrigen Fehlerquote und dem hohen Phi-Koeffizienten. Damit leistet die logistische Regression die genaueste Vorhersage in Bezug auf die Originaltesturteile. Der Entscheidungsbaum zeigt zwar im Vergleich zu den anderen beiden Methoden nur durchschnittliche Gütekriterien, bietet aber durch die leicht interpretierbaren Entscheidungsregeln eine sehr gute Möglichkeit, das Zusammenwirken der einzelnen Prädiktoren zu verdeutlichen. Insgesamt kann man infolge der hohen Trefferquoten und Gütekriterien der multivariat- statistischen Modelle von einer korrekten Vorhersage der Originaltesturteile ausgehen.

### **Vergleich der Prozess- mit den multivariat-statistischen Modellen**

Im Vergleich zeigen sich die Modelle Logistische Regression, „Lautes Denken“ und Diskriminanzanalyse in ihrer Klassifikationsgüte den anderen Modellen überlegen. Die Hypothese, dass ein Prozessmodell durch die Kombination verschiedener Entscheidungsalgorithmen eine quantitativ bessere Vorhersage erbringt, kann aber nicht bestätigt werden. Die Ergebnisse sprechen nur dafür, dass Prozessmodelle hilfreiche Werkzeuge zur Explikation von Entscheidungsregeln sind und besonders die Methode des „Lauten Denkens“ reliable und valide Einblicke in den Entscheidungsprozess liefert. Die Vorhersagegenauigkeit der logistischen Regression wird jedoch von keinem Prozessmodell erreicht.

### **Diskussion**

Einer der Hauptvorteile, der sich durch die Erstellung von Prozessmodellen ergibt und durch die Ergebnisse der Studie bestätigt wird, ist die Prüfung auf Homogenität der Expertenurteile in der gleichen diagnostischen Situation. Der Einsatz von Prozessmodellen ist daher besonders sinnvoll für die Erforschung der Entscheidungsstrukturen einzelner Experten und zur Erstellung von fairen Bewertungsmaßstäben, die im Sinne einer formativen Evaluation als Basis für Validierungsstudien genutzt werden können. Außerdem kann so die Basis zur Diskussion über Qualitätssteigerung, kontinuierliche Qualitätskontrolle und die Möglichkeit des Austausches und Einarbeitung von neuen Mitarbeitern geschaffen werden. Besonders interessant erscheint hier die Bewerberklassifikation durch voneinander unabhängige Entscheidungsregeln die in die Gesamtentscheidung einfließen.

Ein gewichtiger Nachteil dieser Methodengruppe liegt jedoch im hohen Zeitaufwand der für die Explikation aufgebracht werden muss und den damit verbundenen Kosten im Vergleich zu den multiva-

riat-statistischen Ansätzen. Diese zeigen jedoch modellbedingte Nachteile, wenn die Zielsetzung darin besteht, Entscheidungsregeln, transparent abzuleiten und die Befunde aus der Analyse zu interpretieren. Sie tragen daher wenig zum Verständnis des Urteilsprozesses, allerdings stark zur Verbesserung der Urteilsgenauigkeit bei.

Die Wahl eines geeigneten Explikationsmodells hängt damit in hohem Maße von der diagnostischen Zielsetzung, den spezifischen Interessen der Experten und späteren Einsatzbereichen der Modelle ab. Weiterhin sollten für eine möglichst umfassende deskriptive Erfassung von impliziten Entscheidungsregeln, die eine hohe Übereinstimmung mit den Originalurteilen zeigen und interpretierbare Regeln erzeugen, verschiedene Methoden eingesetzt werden. Idealerweise sollten dabei mehrere Ansätze gewählt werden, die sowohl den Process-Tracing wie auch den multivariat-statistischen Ansatz abdecken.

## Literatur

Bono, J. E. & Judge, T.A. (2003). Self-concordance at work: Toward understanding the motivational effects of transformational leaders. *Academy of Management Journal*, 46, 554-571.

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, 87, 137-154.

Büssing, A., Herbig, B. & Latzel, A. (2004). Explikation impliziten Wissens – Verändert sich das Handeln? *Zeitschrift für Psychologie*, 212, 87–106.

Davis, J.H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, 80, 97-125.

Harte, J. & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: Structural modelling, process tracing, and self-reports. *Organizational Behavior and Human Decision Processes*, 64, 49-64.

Höft, S. (2006): Erfolgsüberprüfung personalpsychologischer Arbeit. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*, (2. Auflage, S. 761-796). Göttingen: Hogrefe.

NAGD -Normenausschuss Gebrauchstauglichkeit und Dienstleistungen (2002). *DIN 33430 - Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.

Noack, H; Petermann, F. (1999). Entscheidungstheorie. In R. S. Jäger und F. Petermann (Hrsg.), *Psychologische Diagnostik* (4. Auflage, S. 295-310). Weinheim: Verlags Union.

Ranyard, R. & Williamson, J. (2005) Conversation-based process tracing methods for naturalistic decision making: Information search and verbal protocol analysis. In Montgomery, H.; Lipshitz, R.; Brehmer, B. (Eds.), *How Professionals make Decisions* (pp. 305-318). New Jersey: Lawrence Erlbaum.

Rose, H. (1992). Erfahrungsgeleitete Arbeit als Fokus für Arbeitsgestaltung und Technikentwicklung. *Zeitschrift für Arbeits- und Organisationspsychologie*, 45, 20-28.

Westhoff, K., Hellfritsch, L. J., Hornke, L. F., Kubinger, K. D., Lang, F., Moosbrugger, H., Püschel, A., Reiman, G. (Hrsg.) (2006). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*. Zagreb: Pabst.

Wottawa, H. (1984a). HYPAG: Ein neuer Ansatz in der Marktforschung. *Planung und Analyse*, 11, 15-21.