Moldzio, Thomas; Meyer, Thomas; Glug, Inga & Gulba, Astrid

Implementierung und Evaluation einer praxisorientierten Strategie zur Vorauswahl von Auszubildenden

1. Problemstellung/ Hintergrund

Der Marktführer im Bereich der Hausgerätehersteller stellt in jedem Jahr 40 bis 50 junge Menschen ein, die verschiedene kaufmännische Ausbildungsgänge absolvieren.
Die Personalverantwortlichen stellten fest, dass bis dato zum Bewerbertag (Gruppenübung und Interview) immer wieder Kandidaten eingeladen wurden, bei denen sich schnell herausstellte, dass sie nicht für einen Ausbildungsplatz in Frage kamen. Der Weg zum Bewerbertag stand den Kandidaten offen, die zunächst aus ca. 600 Bewerbungen von der Papierform her selektiert wurden und dann an einem unternehmensintern entwickelten Testverfahren teilgenommen haben. Im Jahre 2001 nahmen z.B. 272 Bewerber am Testverfahren teil. Letztlich erhielten auf der Basis der Testempfehlung 123 Kandidaten eine Einladung zum Bewerbertag, von denen dann 43 eingestellt wurden.
Vor dem Hintergrund dieser Verlaufszahlen und der Feststellung, dass die Qualität der Bewerber nicht mehr der der Vorjahre entsprach, wurde ein externes Beratungsunternehmen beauftragt, den Auswahlprozess im Jahre 2002 für den Ausbildungsjahrgang 2003 zu unterstützen und eine neue Testbatterie zur Vorauswahl zu etablieren. Ziel war es, das Verhältnis zwischen Einladungen zum Bewerbertag und Einstellungen zu optimieren und dadurch parallel eine Kostenreduzierung zu bewirken.

2. Voruntersuchung

Bevor der eigentliche Auswahlprozess begann, wurde eine unternehmensinterne Voruntersuchung durchgeführt, um Anforderungskriterien und entsprechende cut-off-Werte für die Annahme bzw. Ablehnung von Bewerbern generieren zu können. Aus verschiedenen Gründen gab es bisher keine differenzierten Anforderungsprofile, so dass in der Vergangenheit überwiegend erfahrungsgeleitet vorgegangen wurde, d.h. nicht nach expliziten Regeln. Auf der Basis einer Vorgesetzteneinschätzung wurde ein Extremgruppenvergleich realisiert. Hierfür wurden 32 Auszubildende der letzten Jahrgänge – in anonymisierter Form – in eine eher leistungsstarke (N = 21) und eine eher leistungsschwache Gruppe (N = 11) eingeteilt. Diese 32 Auszubildenden bearbeiteten dann eine Testbatterie mit Auszügen aus sieben Leistungstests und Persönlichkeitsfragebögen. Es handelt sich hierbei um die Skalen 3, 4 und 8 des *Leistungsprüfsystems* (LPS) von Horn (1983), die Aufgaben zur Überprüfung der mathematisch-rechnerischen Fähigkeiten aus dem *Intelligenz-Struktur-Test 2000* (I-S-T 2000) von Amthauer, Brocke, Liepmann und Beauducel (1999) und dem *Rechtschreibtest* (RST) von Bullheller und Häcker (2001). Des Weiteren wurden eingesetzt das *NEO-Fünf-Faktoren Inventar* (NEO-FFI) von Borkenau und Ostendorf (1993), das *Arbeitsbezogene Verhaltens- und Erlebensmuster* (AVEM) von Schaarschmidt und Fischer (1996) sowie die Kurzform des *Hamburger Persönlichkeits-Inventars für sechs Basisdimensionen* (HPI-6BD) von Andresen (2002) und ein nicht veröffentlichter Fragebogen zur *Persönlichkeitsstärke*, der auf einer Untersuchung von Hansen und Witt (1984) basiert.
Als Hypothese wurde verfolgt, dass sich die leistungsstarken Auszubildenden hinsichtlich ihrer Ergebnisse in einzelnen kognitiven Leistungsbereichen und in verschiedenen Persönlichkeitseigenschaften von den eher leistungsschwachen unterscheiden. In praxisrelevanten Bereichen, wie der Rechtschreibleistung und in der Gewissenhaftigkeit konnten signifikante Mittelwertunterschiede zwischen beiden Gruppen festgestellt werden. In anderen Bereichen konnten hypothesenkonforme Tendenzen ermittelt werden, die allerdings nicht statistisch signifikant geworden sind, wie z.B. im logisch-schlussfolgernden Denken.
Auf der Basis dieser Voruntersuchung wurden der LPS (3, 4 und 8), der I-S-T 2000 (RA, ZR und RZ), der RST, der NEO-FFI und der AVEM für das eigentliche Vorauswahlverfahren ausgewählt.

3. Hauptuntersuchung

Die Hauptuntersuchung beschäftigte sich mit der Vorauswahl der Bewerber um Ausbildungsplätze für das Jahr 2003. Dabei wurde angenommen, leistungsstarke Bewerber seien besser im logisch-schlussfolgernden Denken, in der Rechtschreibleistung und im Bereich der mathematisch-rechnerischen Fähigkeiten als leistungsschwache Bewerber. Diese Annahme wird gestützt durch die Hinweise aus der Voruntersuchung und die metaanalytischen Untersuchungen von Schmidt und Hunter (1998), die besagen, die allgemeine mentale Fähigkeit sei der beste Prädiktor für Berufserfolg. Des Weiteren wurde angenommen, dass die leistungsstarken Bewerber auf der Ebene der Persönlichkeitseigenschaften über niedriger ausgeprägte Neurotizismuswerte und über höher ausgeprägte Gewissenhaftigkeitswerte verfügen (Howard & Howard, 2002; Moldzio, 1999; Schuler & Barthelme, 1995; Schuler & Funke, 1995). Auf der Basis der qualitativen Eindrücke der Personalverantwortlichen des Unternehmens wurde angenommen, dass kein positiver Zusammenhang zwischen Schulleistungen und Testergebnissen vorliegt. Die Ergebnisse des AVEM wurden nicht weiter verfolgt, da sie keinen differenzierenden Beitrag für den Auswahlprozess geleistet haben.

Zunächst nahmen im August und September 2002 191 Bewerber am Testverfahren zur Vorauswahl teil. Die Bewerber wurden in sechs Gruppen zwischen sieben und 46 Teilnehmern getestet. Instruiert wurden die Testverfahren von einer geschulten Mitarbeiterin des einstellenden Unternehmens. Ausgewertet und interpretiert wurden die Testverfahren unter Leitung eines Diplom-Psychologen des externen Beratungsunternehmens. Die Ergebnisse wurden in Form eines Kurzberichtes über 2,5 Seiten zunächst dargestellt und dann in einer Handlungsempfehlung zusammengefasst. Dabei wurden die Bewerber qualitativ innerhalb einer fünffach abgestuften Beurteilung eingruppiert: A = Potenzialträger (N = 14), B+ = einladen (N = 76), B = eventuell einladen (N = 55), B- = nicht einladen (N = 23), C = auf keinen Fall einladen (N = 23). Für die statistische Auswertung war es vielfach hilfreich, Kombinationen von Eingruppierungen zu betrachten. So wurden die Eingruppierungen A und B+ sowie B- und C häufig zusammengefasst, da es hierbei jeweils um die Hauptzielgruppen der eingeladenen bzw. abgelehnten Bewerber ging. Es handelte sich bei 10% der Bewerber um Realschüler, 44% der Bewerber waren Gymnasiasten und bei 46% handelte es sich um Besucher bzw. Absolventen der Höheren Handelsschule. Der Anteil der Frauen lag bei 70% und der Männeranteil entsprechend bei 30%. Von den Bewerbern waren 15 zwischen 15 und 16 Jahre alt, 115 waren zwischen 17 und 18 Jahre alt, 52 zwischen 19 und 20, während 9 Bewerber zwischen 21 und 24 Jahre alt waren.

Betrachtet man die Ergebnisse der Leistungstests, so konnte zunächst allgemein folgende Abstufung der Gruppenmittelwerte durchgängig registriert werden: M (A, B+) > M (B) > M (B-, C).

Fasst man die Skalen 3 und 4 des LPS zusammen, so zeigen die varianzanalytischen Ergebnisse, dass signifikante Leistungsunterschiede zwischen allen Eingruppierungen im logisch-schlussfolgernden Denken festzustellen waren ($p < .01$). Ebenso zeigten sich signifikante Leistungsunterschiede in der Skala 8 des LPS, die originär das räumliche Vorstellungsvermögen erfasst, aber auch logisch-schlussfolgerndes Denken überprüft (Steller & Stürmer, 1984). Hier unterschieden sich die Gruppen A und B+ von den Gruppen B- und C ($p < .01$) sowie die Gruppe B von der Kombination der Gruppen B- und C ($p < .05$).

Die Ergebnisse zur Überprüfung der mathematisch-rechnerischen Fähigkeit mit Hilfe dreier Aufgabengruppen des I-S-T 2000 (RA, ZR, RZ) zeigen bei allen drei Untertests signifikante Leistungsunterschiede bei den selben Gruppenkombinationen ($p < .01$). Hierbei handelt es sich um die Unterschiede von A und B+ (Potenzialträger und einladen) mit der Eingruppierung in B (eventuell einladen) sowie A und B+ (Potenzialträger und einladen) mit der Kombination der Gruppen B- und C (nicht einladen und auf keinen Fall einladen).

Für den Rechtschreibtest waren signifikante Leistungsunterschiede zwischen allen Eingruppierungen festzustellen ($p < .01$).

Innerhalb der Persönlichkeitsmerkmale konnten signifikante Unterschiede in einzelnen Kombinationen von Eingruppierungen ermittelt werden. So beschrieben sich die Bewerber, die in die Gruppen A und B+ eingestuft wurden, emotional belastbarer (Neurotizismus) als die der Gruppe B. Ebenfalls beschrieben sich A und B+ als emotional belastbarer im Vergleich zu der Kombination

aus B- und C. Die gleiche Reihenfolge und Wertigkeit war ebenso für die Gewissenhaftigkeit feststellbar ($p < .01$).
Die Ergebnisse einer Diskriminanzanalyse zeigen, dass bei alleiniger Betrachtung der durchgeführten Leistungstests (LPS, I-S-T 2000, RST) bereits 75,9% der Bewerber wieder in die gleichen Gruppen eingeteilt werden würden. Bei Hinzunahme der Skalen Neurotizismus und Gewissenhaftigkeit des NEO-FFI trat eine Verbesserung der Eingruppierung um 6,8% ein. Das heißt, wenn man allein diese Untertests und Skalen für die Gruppeneinteilung benutzen würde, so würden 82,7% der Fälle ebenso eingeteilt werden, wie sie ursprünglich im Kurzgutachten eingeteilt worden sind.
Die qualitativen Eindrücke der Personalverantwortlichen des Unternehmens, dass kein positiver Zusammenhang zwischen Schulleistungen und Testergebnissen vorlag, wurde weitestgehend nicht bestätigt. Bei der Betrachtung der Schulleistungen wurden die Noten in den Fächern Deutsch, Mathematik, eine Durchschnittsnote der Sprachen, eine Durchschnittsnote der Naturwissenschaften sowie eine Durchschnittsnote aller ausgewählten Zensuren und einer Gesamtdurchschnittsnote berücksichtigt. Exemplarisch seien hier die regressionsanalytischen Ergebnisse zu den Schulnoten und der entsprechenden Eingruppierung auf Basis der Testergebnisse von 191 Kandidaten erwähnt. So zeigte sich ein Zusammenhang mit der entsprechenden Eingruppierung zwischen den Sprachen ($r = .17^*$), der Mathematiknote ($r = .22^{**}$) und der Gesamtdurchschnittsnote ($r = .16^*$) sowie der Durchschnittsnote der ausgewählten Noten ($r = .22^{**}$). Ebenso gab es signifikante Korrelationen zwischen den Sprachen und dem Ergebnis im Rechtschreibtest ($r = -.21^{**}$), den Deutschnoten und dem Rechtschreibtest ($r = -.16^*$), wie auch den Mathematiknoten und den Ergebnissen der Untertests RA ($r = -.21^{**}$) und ZR ($r = -.25^{**}$) des I-S-T 2000. Bedingt durch die Polung der Schulnoten weisen negative Korrelationen auf einen positiven Zusammenhang hin. Ein signifikanter Zusammenhang zwischen den Leistungen in den naturwissenschaftlichen Fächern und den Testergebnissen konnte nicht festgestellt werden. Die Kennzeichnung der Signifikanzniveaus erfolgte folgendermaßen: $^{**}\ p < .01$ und $^{*}\ p < .05$.
Abschließend wurde die Gruppe der eingestellten Bewerber separat untersucht. Allerdings ergaben sich weder bei den Regressionsanalysen noch bei den Korrelationsrechnungen signifikante Ergebnisse.
Insgesamt wurden von den 191 getesteten Bewerbern 85 zum Bewerbertag eingeladen, wovon 37 Bewerber eingestellt wurden. Hinzu kommen fünf Bewerber, die aus verschiedenen Gründen nicht am Test teilnahmen, schon vorselektiert waren und eingestellt wurden. Weitere acht Bewerber, die Test und Bewerbertag absolviert haben, bekamen ein Angebot für einen Ausbildungsplatz, haben dieses jedoch kurzfristig nicht angenommen bzw. wieder abgesagt.

4. Diskussion

Die Ergebnisse der Hauptuntersuchung zeigen, dass mit Hilfe der gewählten Testverfahren eine Einteilung in drei (A und B+, B, B- und C) bzw. fünf Leistungsgruppen (A, B+, B, B-, C) möglich ist. Der differenzierende Beitrag der Leistungstests in Ergänzung mit der Erfassung zentraler Persönlichkeitseigenschaften konnte aufgezeigt werden. Allerdings sollte hier eine mittelfristige Betrachtung des Berufserfolges der Eingestellten verfolgt werden, um insbesondere den wissenschaftlichen Anforderungen gerecht zu werden. Als Kontrollmechanismus der Güte der Vorauswahl kann zum jetzigen Zeitpunkt die Aussage der Personalverantwortlichen herangezogen werden, dass entgegen der Erfahrungen aus der Vergangenheit die Basisrate bzw. Grundquote der Teilnehmer am Bewerbertag deutlich verbessert werden konnte (Taylor & Russell, 1939). An den Bewerbertagen haben – bis auf eine Ausnahme – ausschließlich Bewerber teilgenommen, die auch grundsätzlich für einen Ausbildungsplatz in Frage gekommen wären und lediglich die Frage der persönlichen Passung geklärt werden musste.
In jedem Fall kann der hohe praktische Nutzen einer Vorauswahl mittels der gewählten Testverfahren beschrieben werden. Neben der Entscheidungsgrundlage für eine Einladung zum Bewerbertag waren nach Aussage der Personalverantwortlichen insbesondere die Handlungsempfehlungen hilfreich, um im Interview gezielt Stärken und Entwicklungsbereiche hinterfragen zu können. Des Weiteren konnte durch die Einführung der beschriebenen Vorgehensweise eine Zeit- und Kostener-

sparnis für das Unternehmen realisiert werden. So nahmen 2002 81 Bewerber weniger am Testverfahren und 38 Bewerber weniger am Bewerbertag teil, was für die Bewerbertage eine Reduzierung der Teilnehmerzahl von 31% ausmacht. Für die Teilnahme am Auswahlprozess ist ein Kostensatz von 300,- € pro Bewerber realistisch, wenn sowohl die administrativen Kosten als auch die Kosten für die Durchführung der Bewerbertage berücksichtigt werden, so dass alleine in der Senkung der Teilnehmerzahl am Bewerbertag ein Einsparungspotenzial von 11.400,- € liegt (38 x 300,- €). Bereits diese Teilbetrachtung der Kostenreduzierung führt weitgehend zu einer Amortisierung der Kosten für die alternative Vorgehensweise mit der beschriebenen Testbatterie.

Auffällig war die Einschätzung der Personalverantwortlichen des Unternehmens, dass kein positiver Zusammenhang zwischen den Schulnoten und den Testergebnissen vermutet wurde. Die empirischen Ergebnisse besagen jedoch, dass in verschiedenen Bereichen sehr wohl Zusammenhänge bestehen. Allerdings besteht ein positiver Zusammenhang zwischen den Testergebnissen und den Schulnoten ausschließlich in der gesamten Stichprobe. Da jedoch bei der alleinigen Untersuchung der Eingestellten weder Regressionsanalysen noch Korrelationen signifikant geworden sind, wird deutlich, dass bei den Personalverantwortlichen der Eindruck entstanden ist, es bestünde kein Zusammenhang zwischen Noten und Testleistung. Die Ergebnisse zeigen, dass hier eine differenzierte Betrachtung notwendig ist.

Nach Abschluss des Bewerbungsverfahrens haben alle eingestellten und abgelehnten Bewerber eine schriftliche Rückmeldung der Testergebnisse erhalten. Hierüber wurde versucht einen Zusatznutzen für die Bewerber zu schaffen, sich als Unternehmen von anderen Ausbildungsunternehmen abzusetzen und die Marke des Unternehmens positiv darzustellen. Der Bereich der Zufriedenheitsbefragung – insbesondere auch der abgelehnten Bewerber – bildet eine wichtige Aufgabe zukünftiger Forschung, der wir z.B. zur Zeit durch eine schriftliche Nacherhebung gerecht zu werden versuchen.

Ein weiteres Feld der Auswahloptimierung liegt darin, Absagen von Auszubildenden nach einer Zusage zu reduzieren. Bis zum 01.12.2002 hatten, wie bereits erwähnt, acht Bewerber ihren Ausbildungsplatz wieder frei gegeben. Diese Problematik hat zugenommen, da die Auswahlverfahren im Wettbewerb immer früher beginnen und teilweise Zeiträume von mehr als einem Jahr zwischen der Zusage und dem Ausbildungsbeginn liegen.

Abschließend sei angemerkt, dass wir im eigentlichen Sinne Personalvorauswahl als den Teil des Auswahlprozesses verstehen, bevor ein Bewerber im Unternehmen vorstellig wird. Somit liegt ein weiterer Schwerpunkt der zukünftigen Optimierung der Vorauswahl darin, Bewerber zu testen, ohne einen Besuch im Unternehmen erforderlich werden zu lassen. Selbstverständlich müsste für einen solch innovativen Schritt u.a. auch die technische Realisierung für alle Beteiligten ermöglicht werden.

Literatur

Amthauer, R.; Brocke, B.; Liepmann, D. & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000.* Göttingen: Hogrefe.

Andresen, B. (2002). *Hamburger Persönlichkeits-Inventar für sechs Basisdimensionen (HPI-6BD).* Göttingen: Hogrefe.

Bullheller, S. & Häcker, H.O. (2001). *Rechtschreibtest.* Frankfurt: Swets Test Services.

Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI).* Göttingen: Hogrefe.

Hansen, J. & Witt, H.-J. (1984). Persönlichkeitsstärke. *Planung und Analyse*, *11* (3) ,112-117.

Horn, W. (1983). *Leistungsprüfsystem. L-P-S* (2. erw. Aufl.). Göttingen: Hogrefe.

Howard, P. & Howard, J.M. (2002). *Führen mit dem Big-Five Persönlichkeitsmodell.* Frankfurt: Campus Verlag.

Moldzio, T. (1999). *Prädiktoren für den Personalvorauswahlprozess im Vertriebsbereich. Erfolgreiche Repräsentanten der Finanzdienstleistung – eine besondere Spezies?* Unveröffentlichte Diplomarbeit. Universität Kiel.

Schaarschmidt, U. & Fischer, A. (1996). *Arbeitsbezogenes Verhaltens- und Erlebensmuster.* Frankfurt: Swets Test Services.

Schmidt, F.L. & Hunter, J.E. (1998). Meßbare Personenmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauß (Hrsg.), *Potentialfeststellung und Personalentwicklung* (15-43). Göttingen: Verlag für Angewandte Psychologie.

Schuler, H. & Barthelme, D. (1995). Soziale Kompetenz als berufliche Anforderung. In B. Seyfried (Hrsg.), *"Stolperstein" Sozialkompetenz* (77-116). Bielefeld: Bertelsmann.
Schuler, H. & Funke, U. (1995). Diagnose beruflicher Eignung und Leistung. In H. Schuler, *Lehrbuch Organisationspsychologie* (2., korr. Aufl, 235-283). Bern: Verlag Hans Huber.
Steller, M. & Stürmer, M. (1984). Space, Reasoning oder was? Zur Validität des "Raumvorstellungstests" LPS 8/PSB 7. *Zeitschrift für experimentelle und angewandte Psychologie, XXXI*, 671-694.
Taylor, H.C. & Russell, J.F. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

Erscheint im Frühjahr 2003 im Processing-Band des 9. Dresdner Symposiums "Psychologie der Arbeit" (27.02.-01.03.2003) im Rahmen des 175-jährigen Jubiläums der Technischen Universität Dresden.