

DAS AUSWAHLGESPRÄCH BEIM HOCHSCHULZUGANG - ERFahrungen DER STUDIENSTIFTUNG UND ERGEBNISSE DER INTERNATIONALEN INTERVIEWFORSCHUNG

G. Trost, Bonn

1. Erfahrungen der Studienstiftung beim Auswahlverfahren für Oberprimaner

In den Jahren 1970 bis 1980 führte die Studienstiftung des deutschen Volkes ein dreistufiges Auswahlverfahren für Bewerber um ein Stipendium durch, welches bereits vor dem Ablegen der Reifeprüfung abgeschlossen war und die Möglichkeit bot, die aufgenommenen Kandidaten vom ersten Tag ihres Studiums an zu fördern. Sechs Bundesländer waren in dieses Auswahlprogramm einbezogen. (Seit 1981 kann dieses Verfahren aus finanziellen Gründen nicht mehr durchgeführt werden.)

In der ersten Auswahlstufe wurden alle Gymnasien gebeten, auf der Grundlage des Notendurchschnitts in den Versetzungszeugnissen nach dem 12. Schuljahr eine Spitzengruppe von 10 Prozent unter den Schülern des anstehenden Abiturjahrgangs zu nominieren. Alle nominierten Schüler wurden zur Teilnahme an einer schriftlichen Prüfung, dem "Auswahltest der Studienstiftung" (ATS), eingeladen. Der ATS ist ein vom Institut für Test- und Begabungsforschung der Studienstiftung nach amerikanischen Vorbildern entwickelter allgemeiner Studierfähigkeits-test, dessen Schwierigkeitsgrad auf das hohe Leistungsniveau der notenbesten Oberprimaner zugeschnitten ist. Der Test besteht aus einem Verbalteil und einem quantitativen Teil und dauert etwas über

drei Stunden; er wurde in allen an diesem Auswahlprogramm beteiligten Bundesländern am selben Tag durchgeführt. Aufgrund der Leistungen im Auswahltest wurden etwa 30 Prozent der nominierten Kandidaten als "Finalisten" zur dritten Auswahlstufe zugelassen. Dabei handelte es sich um dreitägige Auswahlseminare, in deren Verlauf jeder Bewerber

- (a) zwei Einzelgespräche mit zwei verschiedenen Mitgliedern der Auswahlkommission führte,
- (b) an neun Gruppensitzungen zusammen mit acht Mitbewerbern teilnahm, in denen jeweils ein Bewerber ein Referat mit einem Thema eigener Wahl hielt, über welches im Anschluß in der Gruppe diskutiert wurde,
- (c) sich an einer "führerlosen Gruppendiskussion" beteiligte; während einer Stunde war in einer Kleingruppe von vier bzw. fünf Teilnehmern ein Thema zu bearbeiten und eine Gruppenlösung zu finden.

Jedes Mitglied der Auswahlkommission hatte 36 Bewerber zu beurteilen, davon 18 aufgrund der Beobachtungen beim Einzelgespräch, 9 aufgrund der Leistungen bei den Gruppendiskussionen mit Referaten und 9 aufgrund des Verhaltens bei den führerlosen Gruppen-

diskussionen. Umgekehrt traf jeder Bewerber mit vier Mitgliedern der Auswahlkommission zusammen. Während der zehn Jahre, in denen dieses Auswahlprogramm durchgeführt wurde, nahmen über 14.000 Stipendienbewerber an Auswahlseminaren teil.

Die Einzelgespräche (Interviews) wurden vorbereitet durch einen Fragebogen "Biographische Daten", durch welchen der Interviewer Informationen über den biographischen Hintergrund, die familiäre Situation, schulische und außerschulische Interessen und Aktivitäten, Studien- und Berufspläne sowie besondere Erfahrungen des Bewerbers erhielt. Der Fragebogen diente zugleich als Leitfaden für das Auswahlgespräch. Jedes Auswahlgespräch dauerte 45 Minuten; zwischen je zwei Gesprächen lagen 15 Minuten für Vor- und Nachbereitung.

Am Vorabend der Auswahlseminare wurden die Mitglieder der Auswahlkommission vom Referenten der Studienstiftung über die folgenden Festlegungen zur Bewertung der Auswahlgespräche instruiert:

- Das Gesamturteil über die Eignung eines Bewerbers für ein Stipendium sollte anhand einer zehnstufigen Skala abgegeben werden.
- Bei der Vergabe der Punkte sollte sich der Prüfer an der Qualifikation aller beobachteten Bewerber orientieren; die Skala sollte voll ausgeschöpft werden; der Mittelwert der Urteile sollte bei etwa 5,5 Punkten liegen.
- Die Testwerte der Bewerber waren den

- Prüfern nicht bekannt, wohl aber die Schulnoten, die im Fragebogen gemachten Angaben und ein kurzer Lebenslauf.
- Wertungen über Einzelpersonen durften zwischen den Mitgliedern der Auswahlkommission vor der Abschlusssitzung nicht ausgetauscht werden.
- In der Abschlusssitzung wurden alle Einzelwertungen zusammengetragen; stark abweichende Wertungen über denselben Kandidaten wurden stets diskutiert.

Aufgrund der Summe aus allen Einzelwertungen im Auswahlseminar und einem Viertel des Test-Gesamtwertes wurden jeweils alle Teilnehmer an einem Auswahlseminar (54 Kandidaten) in eine Rangfolge gebracht. Ein Drittel bis ein Viertel der Kandidaten konnte in die Studienstiftung aufgenommen werden. In der Festlegung der "Schnittstelle" waren die Auswahlkommissionen frei. Der Schwerpunkt der Diskussionen bei der Abschlusssitzung lag deshalb auf der Erörterung der "Grenzfälle". Die Entscheidungen über die Aufnahme wurden in der Regel einstimmig von der gesamten sechsköpfigen Auswahlkommission gefällt.

Den Mitgliedern der Auswahlkommission stand außer den in der Satzung der Studienstiftung festgelegten allgemeinen Auswahlkriterien für Stipendienbewerber ("... fördert die Hochschulbildung junger Menschen, deren hohe wissenschaftliche oder künstlerische Begabung und deren Persönlichkeit besondere Leistungen im Dienst der Allgemeinheit

erwarten lassen...") keine differenzierte Liste von Eignungsmerkmalen zur Verfügung. Eine gewisse Pluralität in der Gewichtung und Bewertung von Einzelinformationen über die Kandidaten war erwünscht.

In einer Begleituntersuchung an einem ganzen Bewerberjahrgang (858 Interviews, 102 Interviewer) konnte die Übereinstimmung der Urteile zwischen jeweils zwei Interviewern über denselben Kandidaten aufgrund getrennter Einzelgespräche ermittelt werden. Die Rangkorrelation zwischen den Urteilen betrug $r_{tt} = .55$. Dieser Kennwert für die Beurteiler-Übereinstimmung ist zwar deutlich niedriger als die Kennwerte für die Objektivität von Tests; er ist indessen keineswegs niedriger als die vorliegenden Kennwerte zur Objektivität der Zensurengebung in der Schule. Der Kennwert für die Korrelation der Einzelgespräche mit dem Ergebnis im Auswahltest betrug $r = .21$, der Korrelationswert für die Beziehung zum Notendurchschnitt im Gymnasium $r = .25$. Beide Koeffizienten lassen erkennen, daß in den Einzelgesprächen weit überwiegend solche Merkmale der Bewerber bewertet wurden, die nicht bereits in der schulischen Leistung oder in der Testleistung zum Ausdruck gekommen waren.

Bei einer Längsschnitt-Bewährungskontrolle zum Auswahlverfahren, in die nicht nur aufgenommene, sondern auch abgewiesene Stipendienbewerber einbezogen werden konnten und die sich auf Studienverlaufsdaten von 1.906 Studierenden bezog (LAAGLAND 1978), konnte u.a. der Studienerfolg von Bewerbern, die bei den Auswahlseminaren erfolgreich gewesen waren ($N = 129$), mit dem Studienerfolg jener verglichen werden, die beim Auswahlseminar abgelehnt worden waren ($N = 272$). Nach einem komplexeren Kriterium für den Studienerfolg, in welches neben den Examensnoten auch eine allgemeine Selbsteinschätzung bezüglich der Studienleistungen einging, konnte für die im Auswahlseminar nicht erfolgreichen Studierenden ein hoher Studienerfolg in 80 Prozent der Fälle ermittelt werden; bei den im Auswahlseminar erfolgreichen Studierenden war ein hoher Studienerfolg in 92 Prozent der Fälle zu verzeichnen. Da die beiden Einzelgespräche wichtigste Grundlage für die Aufnahmeentscheidung bei den Auswahlseminaren waren, läßt sich aus diesem Ergebnis, welches noch dazu an einer bereits extrem ausgelesenen Gruppe (etwa 3 Prozent des entsprechenden Abiturjahrgangs) ermittelt worden ist, auf eine gewisse Prognosekraft des Interviews schließen.

2. Übersicht über Ergebnisse der internationalen Interviewforschung

Ein Überblick über die Ergebnisse von Untersuchungen zum Interview bei der

Vorhersage von Leistungen im tertiären Bildungsbereich aus verschiedenen

Ländern führt zu den folgenden Schlüssen (vgl. TROST 1982, 1984).

Das Interview ist das mit Abstand am meisten verwendete Auswahlverfahren rund um die Welt. In eklatantem Gegensatz zur Häufigkeit der Verwendung steht die Anzahl der (veröffentlichten) Untersu-

chungen zur Güte und Bewährung des Interviews. Klassische Kriterien zur Bewertung eines diagnostischen Verfahrens sind die der "Reliabilität" (Zuverlässigkeit, Meßgenauigkeit) und der "Validität" (Gültigkeit, Prognosekraft). Unter diesen beiden Aspekten seien die zugänglichen Forschungsergebnisse resümiert.

(a) Reliabilität des Interviews

Sie läßt sich definieren als die Stabilität der Urteile eines Interviewers, wenn er dieselben Kandidaten mehrmals interviewt; dieser Aspekt ist bisher kaum untersucht worden. Eine weitere Definition bestimmt die Reliabilität als die Übereinstimmung der Urteile zweier oder mehrerer Interviewer über dieselben Kandidaten; zu diesem Aspekt liegen zahlreiche Untersuchungsergebnisse vor. Sie weisen auf eine bestenfalls mäßige Übereinstimmung der Interviewer-Urteile über dieselben Kandidaten hin. Die Kennwerte für das Ausmaß dieser Übereinstimmung sind stets deutlich niedriger, als sie üblicherweise für Instrumente der Diagnostik, insbesondere in Auslesesituationen, gefordert werden.

Bei näherer Betrachtung der einzelnen Untersuchungen zeigen sich Hinweise darauf, welche Bedingungen in besonderem Maße die Zuverlässigkeit der Interview-Ergebnisse beeinträchtigen:

- Eigenschaften des Interviewers: die Milde oder Strenge seines Urteils;

sein eigenes verbales oder nonverbales Verhalten, das die Reaktionen des Gesprächspartners beeinflußt; seine eigenen Einstellungen, Wertvorstellungen und Interessen, die Einfluß auf die Bewertung der entsprechenden Äußerungen des Interviewten nehmen; schließlich das Ausmaß an Training und Erfahrung in der Führung von Auswahlgesprächen.

- Eigenschaften des Interviewten, die für den eigentlichen diagnostischen Zweck des Interviews irrelevant sind: seine äußere Erscheinung, Mimik, Gestik usw.
- Besonderheiten der Interviewsituation: die affektive Beziehung zwischen Prüfer und Bewerber ("Rapport"); Verlauf und Ergebnisse der vorangegangenen Interviews (z.B. Kontrasteffekte); Umfang und Art der Vorinformation des Interviewers über den Kandidaten; die Abfolge, in der der Interviewer Informationen während des Interviews erhält.

Den empirischen Befunden zufolge

dürfen zuverlässigere Interview-Ergebnisse erwartet werden, wenn

- das Interview strukturiert ist,
- die Kriterien der Beurteilung genau umrissen sind,
- die Urteile über die Bewerber anhand definierter Kategorien oder Skalen gefällt werden,
- die Interviewer sorgfältig ausgewählt und trainiert sind,
- mehr als ein Interviewer pro Kandidat

zur Verfügung steht und

- jeder Interviewer eine genügend große Zahl von Bewerbern kennenlernt.

Auch wenn diese und weitere formale Vorgaben durchgesetzt werden können, wird das Auswahlgespräch nie die Objektivität eines standardisierten Tests erreichen, wenn es mehr bleiben soll als ein reines, vorformuliertes Frage-und-Antwort-Spiel, das leicht durch einen Fragebogen ersetzt werden könnte.

(b) Validität (Gültigkeit des Interviews)

Im Zusammenhang mit der Verwendung bei Auswahlentscheidungen läßt sich die Validität des Interviews am besten definieren als die Beziehung zwischen dem Interviewergebnis und der späteren Leistung des Beurteilten im Ausbildungsabschnitt, zu dem er zugelassen werden soll, mithin als die Vorhersagekraft des Interviews.

Die vorliegenden Ergebnisse der Interviewforschung zu dieser Fragestellung sind extrem uneinheitlich; die Kennwerte für die Prognosekraft der Interviewurteile sind über einen breiten Streubereich gespannt. Dies überrascht nicht angesichts der erheblichen Unterschiede in den Zielen, Inhalten und Abläufen der Interviews in den jeweiligen Untersuchungen.

Eine Beurteilung der genuinen Prognosekraft des Interviews wird dadurch erschwert, daß meist unklar ist, ob die Interviewurteile allein auf Beobach-

tungen während des Interviews oder auch auf der Kenntnis bisheriger Leistungen der Kandidaten, also etwa ihrer Schulnoten und ihrer Testergebnisse, beruhen.

Zusammenfassend läßt sich allerdings feststellen, daß die große Mehrzahl der vorliegenden Untersuchungsergebnisse auf eine sehr geringe Vorhersagekraft des Interviews hinweist. Nimmt man das Interview zu anderen Vorhersageinstrumenten hinzu, so ist der Zugewinn an Prognosekraft ebenfalls gering. Zwar unterschätzt man die Prognosekraft eines Ausleseinstruments stets dann, wenn man nur die zugelassenen Kandidaten in die Bewährungskontrollen einbeziehen kann - denn die Kandidaten mit ungünstigen Interviewergebnissen, die vermutlich auch künftig schlechtere Leistungen vorzuweisen hätten, müssen unberücksichtigt bleiben. Selbst wenn man diesen Umstand in Rechnung stellt, erweisen sich unter vergleichbaren Bedingungen aber Schulnoten und Testwerte

durchgängig als die besseren Vorhersagefaktoren.

3. Folgerungen und Empfehlungen

- Nichts, was durch einen guten Test gemessen werden kann, sollte dem Interview überlassen bleiben.
 - Es gibt indessen wichtige Eigenschaften, die besser durch das Interview als durch andere Verfahren geprüft werden können; allein das Interview erlaubt etwa ein gezieltes, individuelles Nachfragen bei unbefriedigenden und uneindeutigen Äußerungen des Kandidaten, ferner eine ganzheitliche Beobachtung und Würdigung von verbalen Äußerungen, Auftreten und Ausdrucksverhalten sowie des Sozialverhaltens in einer schwierigen Gesprächssituation. Auf diese Merkmale, die beim Auswahlgespräch beobachtbar sind, sollte sich die Bewertung auch konzentrieren, also z.B. auf das Kommunikations-Verhalten, die Fähigkeit, sich auf einen Gesprächspartner einzustellen, auf die Darlegung und Begründung von Interessen, Einstellungen und Motiven, die Art und Weise der Verarbeitung bisheriger studien- und berufsbezogener Erfahrungen, die Überzeugungskraft und Begründetheit der Äußerung von Meinungen, die Flexibilität im Eingehen auf wechselnde Gesprächsgegenstände. Fakten sollten im Interview nicht abgefragt werden; dafür eignet sich ein vorgeschalteter, standardisierter Fragebogen besser.
 - Die Erwartungen in die Prognosekraft des Interviews bezüglich späterer Leistungen in Studium und Beruf dürfen nicht hoch gespannt werden.
 - Dagegen hat das Interview andere Vorzüge gegenüber den verfügbaren standardisierten diagnostischen Verfahren: Es ermöglicht dem Bewerber die persönliche Begegnung mit denen, die über seine Bewerbung entscheiden, und den Kontakt mit der auswählenden Institution; es bietet die Gelegenheit, die Hochschule in die Verantwortung für die Auswahl des studentischen Nachwuchses einzubinden; es erlaubt die Berücksichtigung individueller Lebensumstände und Einzelschicksale bei der Bewertung bisheriger Aktivitäten und Leistungen; es ermöglicht eine Rückmeldung an den Kandidaten, z.B. auch im Sinne einer Darstellung der Bildungsziele und der Anforderungen der jeweiligen Hochschule.
- Für die Durchführung und Bewertung des Auswahlgesprächs bei der Zulassung zu den medizinischen Studiengängen wird folgendes empfohlen:
- Eine Strukturierung oder "Teilstandardisierung" scheint notwendig; sie erhöht die Zuverlässigkeit der Interviewergebnisse, wahrscheinlich auch deren Prognosekraft, ferner die "Gerichtsfestigkeit" des Verfahrens, indem sie ausschließt, daß die Abläufe

- der Interviews, die angesprochenen Themenbereiche und die Bewertungskriterien von Prüfer zu Prüfer völlig verschieden sind.
- Die Dauer des Interviews sollte einheitlich festgelegt werden; sie sollte nicht weniger als eine halbe Stunde betragen.
 - Die Anzahl der Bewerber pro Interviewer sollte annähernd einheitlich sein.
 - Die Vorinformationen der Prüfer sollten einheitlich sein: Sie werden am besten durch einen Fragebogen erhoben, der Fakten und Einschätzungen erfragt. Auch die Fachnoten in der Schule sollte der Prüfer kennen. Ob ihm die Ergebnisse im "Test für medizinische Studiengänge" bekannt sein müssen, sollte eingehend geprüft werden.
 - Der Fragebogen sollte zugleich als Leitfaden für den Gesprächsverlauf dienen; er sollte die Themenbereiche abstecken. Der Prüfer sollte die vom Kandidaten im Fragebogen gemachten Angaben "hinterfragen", nach dem Warum und dem Wie forschen, die Ernsthaftigkeit und Intensität der aufgeführten Aktivitäten und Motive zu erfassen versuchen.
 - Es sollten einige wenige allgemeine Kriterienbereiche definiert werden, die bei der Gesamtwertung des Interviewergebnisses berücksichtigt werden sollen. Die Gewichtung der Einzelkriterien für das Gesamturteil sollte dem Prüfer überlassen bleiben.
 - Das Gesamturteil des Prüfers sollte auf einer einfachen Punkteskala abgegeben werden; Einstufungen nach mehreren Kriterien auf gesonderten Skalen werden nicht empfohlen.
 - Eventuell sollten einige zentrale Begründungen für das Gesamturteil in wenigen Stichworten festgehalten werden.
 - Zwei Einzelinterviews sind diagnostisch ergiebiger als ein Interview mit zwei Prüfern, allerdings ist die Belastung des Prüfers beim Einzelinterview höher.
 - Zur Vermeidung von Häufungen der Urteile im positiven oder negativen Bereich sollten die Prüfer angehalten werden, die gesamte verfügbare Skala zur Bewertung der von ihnen beobachteten Kandidaten auszunutzen.
 - Zur Entscheidung einer Hochschule über Aufnahme oder Ablehnung der ihr zugewiesenen Interview-Teilnehmer sollte anhand der Summe der Wertungen für jeden Kandidaten durch alle beteiligten Prüfer eine Rangfolge über alle an der Hochschule interviewten Kandidaten erstellt werden; in einer Schlußbesprechung, an der alle Prüfer beteiligt sind, sollten deutlich diskrepante Wertungen diskutiert werden und durch den Austausch von Informationen ggfs. Veränderungen der Rangpositionen einzelner Kandidaten vorgenommen werden. Insbesondere im Bereich der Schnittstelle sollte jeder Kandidat besprochen werden. Die endgültige Entscheidung über die aufgenommenen Kandidaten sollte einstimmig oder, bei großen Prüfungskommissionen, mindestens mit Zweidrittelmehrheit ge-

- fällt werden.
- Alle am Auswahlgespräch beteiligten Prüfer sollten an einer Einweisung teilnehmen, in der Hinweise zum Ablauf und zu den Inhalten des Auswahlge-

sprächs, zu den Bewertungskriterien und Bewertungsverfahren und nicht zuletzt zur Technik der Gesprächsführung in einer für das Bundesgebiet einheitlichen Weise vermittelt werden.

4. Literaturhinweise

LAAGLAND, E. 1978: Evaluierung eines Auswahlverfahrens für die Ermittlung der Studierbefähigung. Nacherhebung zur Oberprimanerauswahl der Studienstiftung des deutschen Volkes. Minerva-Publikation, München.

TROST, G. 1982: Das Interview als Instrument zur Diagnose der Studien-

eignung. In: STUDIENSTIFTUNG (Hg.): Jahresbericht 1981. Bonn, S. 171-185.

TROST, G. 1984: Interviews for Selection. In: HUSÉN, T. & POSTLETHWAITE, T.N. (Eds.): The International Encyclopedia of Education. Research and Studies. Pergamon Press, Oxford, im Druck.

In der Diskussion teilt der Referent mit, daß dieses Auswahlverfahren aus finanziellen Gründen nur in sechs Bundesländern durchgeführt werden konnte (unter anderem war Nordrhein-Westfalen nicht beteiligt). Weiterhin stellt er noch einmal die Bedeutung der führerlosen Gruppendiskussion als hervorragendes Instrument zur Untersuchung des Sozialverhaltens und von Führungseigenschaften heraus.- Angesichts der gründlichen Evaluation des Auswahlverfahrens der Studienstiftung sollte auch bei den Auswahlgesprächen für die Zulassung zum Medizinstudium von Anbeginn auf eine Evaluation gedrungen werden.