

Persönliche und soziale Orientierungslagen von Hochschullehrern/innen der Psychologie zu Evaluationskriterien über eigene berufliche Leistungen

Befunde einer Expertenbefragung bei Professoren/innen und Privat-Dozenten/innen der Psychologischen Institute mit Hauptfachausbildung in der Bundesrepublik Deutschland

Leo Montada, Günter Krampen und Patrick Burkard

Zusammenfassung. Unter Bezug auf die Bestrebungen und Erfordernisse nach einer Evaluation – der Qualitätssicherung und dem Qualitätsmanagement – der Arbeit von Hochschullehrern(innen) und Universitätsinstituten wurde eine Expertenbefragung durchgeführt, um die Frage zu klären, ob für dabei potentiell verwendbare Evaluationskriterien in der Gruppe der an deutschen Psychologischen Instituten mit Hauptfachausbildung tätigen Hochschullehrer(innen) ein kollegialer Konsens existiert. Präsentiert werden Befunde, die in einer Stichprobe von 265 Experten in eigener Sache gewonnen wurden und sich auf die persönlichen Bewertungen von 117 zur Beurteilung vorgegebenen möglichen Evaluationskriterien, die darauf bezogenen sozialen Orientierungslagen und die Konsensfähigkeit der einzelnen Kriterien sowie die Dimensionalität der auf die Kriterien bezogenen Wichtigkeitseinschätzungen beziehen.

Schlüsselwörter: Hochschullehrer, Wissenschaftler, Psychologieausbildung, berufliche Leistung, Evaluation, Leistungsbewertung, Psychologen, Wissenschaftsforschung

Personal and social orientations of psychology college teachers on evaluative criteria for own job performances: Results of an expert survey in German graduate psychology college teachers

Abstract. Presents the results of an expert survey on the importance of criteria for the evaluation of job performance of college teachers in psychology teaching and psychological research. Questionnaire data were obtained in a representative sample of 265 German university teachers working in undergraduate and graduate psychology education. Used was a rather exhaustive list of 117 possible evaluative criteria, which were rated with regard to (a) personal importance and (2) judged importance within the vocational reference group of German psychology college teachers. Combining these ratings, four types of conformity (conscious conformity, supposed conformity, supposed deviance, conscious deviance) and three indices of conformity (knowledge of the social norm, subjective conformity, objective conformity) are derived. The results point towards a socially well integrated vocational group with rather high consensus about the high importance of 100 evaluative criteria and a somewhat less importance of 9 criteria. The large majority of the sample knows the social norms with regard to the importance of the criterias and belongs to the type of conscious conformity, indicating rather valid cognitive representations of group-specific personnel evaluation criteria. Results refer to the personal and supposed importance ratings of all evaluative criteria as well as to their dimensionality and the possibility to develop psychometric subscales for the measurement of the importance of different evaluation dimensions in the context of research productivity and teaching quality of psychology college teachers.

Key words: college teachers, scientists, graduate psychology education, job performance, evaluation, personnel evaluation, vocational evaluation, psychologists, science research

Erfordernisse nach Evaluation oder – in modernem Sprachgebrauch – Qualitätskontrolle, Qualitätssicherung und Qualitätsmanagement werden seit gut einer Dekade zunehmend an Bildungseinrichtungen herangetragen und

– partiell – auch in und von ihnen offensiv propagiert (vgl. hierzu etwa Deutsche Gesellschaft für Psychologie, 1997; Felt, Nowotny & Taschwer, 1995; Heckhausen, 1988; Hellstern, 1997; Mummendey & Wottawa, 1996).

Dies gilt auch für die Universitätsausbildung im Fach Psychologie, die allerdings im universitären Fächervergleich in der Bundesrepublik Deutschland bislang dadurch, daß sie zu den wenigen harten „*numerus clausus*“-Fächern gehört, (noch) in einem gewissen Schonraum verblieben ist: Bei nach wie vor hoher, sogar in den letzten Jahren zugenommener Studienplatznachfrage (siehe hierzu etwa Spada, 1997) werden Studierende den Psychologischen Instituten der verschiedenen deutschen Universitäten von der Zentralstelle für die Vergabe von Studienplätzen (ZVS) zugewiesen. Wahlmöglichkeiten von Studierenden, die etwa auf evaluativen „rankings“ der Institute basieren könnten, spielen allenfalls nach dem Vordiplom eine Rolle (wobei aber auch hier Zulassungsbeschränkungen für höhere Fachsemester regulierend eingreifen). Fachinitiativen zur Evaluation eigener Arbeitsstandards bleiben bislang weitgehend auf die fachinterne wissenschaftstheoretische Diskussion von Gütekriterien für die wissenschaftliche Arbeit (für die in der deutschsprachigen Fachliteratur im Vergleich der 90er mit den 80er Jahren aber auch ein Abflauen zu konstatieren bleibt: Während für die Publikationsjahrgänge 1980 und 1989 noch knapp acht Promille des gesamten im PSYNDEX nachgewiesenen psychologischen Literaturaufkommens mit dem freien Deskriptor „Wissenschaftstheorie“ versehen ist, gilt dies für die Zeit von 1990 bis 1997 nur mehr für knapp vier Promille der psychologischen Fachliteratur) sowie vereinzelte Bemühungen um eine testtheoretische Fundierung der Evaluation der Lehre über Studierendenbefragungen (häufig aber primär bezogen auf Studierende anderer Fächer; vgl. etwa Daniel, 1994; Rindermann, 1998; Rindermann & Amelang, 1994) begrenzt. Allenfalls dann, wenn Psychologische Institute innerhalb einer Universität – etwa bei der Verteilung von Haushaltsmitteln zwischen und/oder innerhalb von Fakultäten (vgl. etwa Deutsche Gesellschaft für Psychologie, 1997; Mummendey & Wottawa, 1996) – oder zwischen verschiedenen Universitäten unter direkten, zumeist (finanz-)politisch initiierten Konkurrenzdruck geraten, wurden hier und da Initiativen zu einer Evaluation von Instituten und damit auch Personen gestartet, die aber oftmals wegen der mangelnden Transparenz der Herkunft von Evaluationskriterien sowie den inhaltlichen und methodischen Problemen bei ihrer Umsetzung umstritten bleiben.

Ähnliches gilt etwa auch für die an Analysen des „*Social Science Citation Index*“ (SSCI) orientierte Evaluation des Publikationsaufkommens deutschsprachiger Psychologen und Psychologinnen nach deutsch- versus englischsprachigen Veröffentlichungen sowie deren Zitationsraten (d. h., den „*impact*“ der deutsch- versus englischsprachigen Publikationen auf die internationale psychologische Forschung) von Keul, Gigerenzer und Stroebe (1993; vgl. auch die erheblich frühere SSCI-Analyse für den angloamerikanischen Bereich auf der Institutsebene von Endler, Rushton & Roediger, 1978). Kritisiert wurde an diesem Beitrag vieles (vgl. zusammenfassend Tack, 1994), einiges war daraufhin zu korrigieren (siehe Keul, Gigerenzer & Stroebe, 1994), anderes blieb bestehen (siehe Tack, 1994), wobei das Grundpro-

blem der (fast prinzipiellen) Schwäche eines singulären Evaluationskriteriums (hier: Zitationen von Zeitschriftbeiträgen nach dem nicht-exhaustiven SSCI, in dem lediglich Erstautoren berücksichtigt werden, deren Namen zudem häufig in unterschiedlichen Schreibweisen dokumentiert sind etc.; vgl. hierzu auch Montada, Becker, Schoepflin & Baltes, 1995) mit unbekannter Akzeptanz bei den von den evaluativen (d. h., bewertenden!) Analysen direkt Betroffenen offen zu Tage getreten ist und nachhaltig blieb.

Wird nun aber der Annahme zugestimmt, daß dem universitätsexternen (etwa von Ministerien, Politikern, Massenmedien und der Öffentlichkeit ausgehenden), dem universitätsinternen (etwa bedingt durch eine leistungsbezogene Mittelvergabe zwischen den Fachbereichen/Fakultäten und zwischen den Fächern; vgl. Deutsche Gesellschaft für Psychologie, 1997; Hellstern, 1997) und dem fachinternen (etwa bedingt durch eine leistungsbezogene Mittelvergabe zwischen verschiedenen Abteilungen, siehe Deutsche Gesellschaft für Psychologie, 1997, und bei Berufungsverhandlungen; siehe hierzu auch Spada, 1997) Evaluationsdruck auch von Psychologischen Instituten ebenso wie im Ausland (vgl. etwa van Raan, 1993, 1996a; im Überblick siehe etwa Felt et al., 1995, S. 235 ff.) in der Zukunft kaum (weiter) ausgewichen werden kann, so stellt sich die Frage nach den dabei anzusetzenden evaluativen Kriterien. Versuche, sich auf ein singuläres Evaluationskriterium zu konzentrieren, werden nicht allein wegen deren prinzipieller inhaltlicher und methodischer Probleme ausscheiden, sondern auch wegen der erheblichen (potentiellen) Breite und Komplexität der Tätigkeit von Hochschullehrern/innen in Psychologischen Instituten. A priori sinnvoller ist eine (möglichst exhaustive) Liste von Evaluationskriterien, bei der – ähnlich wie in der Psychometrie – darauf gesetzt wird, daß Meßfehler nicht (oder zumindest nicht allzu hoch) korreliert sind und sich die Fehlerbehaftetheit der verschiedenen Indikatoren (Items) wechselseitig aufhebt, so daß ein reliables und möglichst auch valides Aggregat entstehen kann. Konstruktvalidität ist dann gewährleistet, wenn die Evaluationskriterien (ggfs. mit unterschiedlichem Gewicht) inhaltlich gültige Indikatoren für das hypothetische Konstrukt der „guten“ Arbeit/Produktivität Psychologischer Institute und der dort tätigen Hochschullehrer(innen) sind. Die „Gültigkeit“ solcher Kriterien kann von außen (etwa von Ministerien, Massenmedien oder der Öffentlichkeit) aufgrund von Plausibilitätsüberlegungen, politischen Zielvorgaben oder Annahmen über die Öffentlichkeits- und Werbewirksamkeit schlicht postuliert werden (man denke etwa an die in den letzten Jahren in politischen Magazinen und Illustrierten publizierten „rankings“ von Universitäten und Studienfächern; zur allgemeinen und fachinternen Kritik daran siehe etwa Daniel, 1986; Daniel, Frühwald, Schiedermaier & Erichsen, 1995; Heckhausen, 1988; Hornbostel & Daniel, 1994; Krampen, Burkard, Wiesenhütter & Montada, 1998; Süllwold, 1997a). Dieses Vorgehen führt in der Regel zu geringer Akzeptanz der Befunde in der Fachöffentlichkeit (bzw. zu höchst unterschiedlicher Akzeptanz, die von heimlicher Freude bei den „Gewinnern“

über Ignoranz bei den Mittelplazierten bis zu massiver Ablehnung aufgrund methodischer Kritik bei den „Verlierern“ variiert, wobei jedoch die methodischen und inhaltlichen Bedenken in der Regel von allen geteilt, jedoch nicht unbedingt auch verbalisiert werden). Als Alternative zu dieser externen, nicht oder zumindest wenig (bzw. manchmal vielmehr allzu) transparenten Setzung von Evaluationskriterien mit unklarer Objektivität, Reliabilität und Validität ist zu prüfen, ob entsprechende Evaluationskriterien für die eigene Tätigkeit aus dem Fach heraus (mithin intern) konsensuell entwickelt werden können. Danach ist ein empirischer Weg der Bestimmung von Evaluationskriterien zu beschreiten, von dem zu erwarten ist, daß er – falls die konsensuelle Validierung von Evaluationsindikatoren gelingt – zugleich eher zu dem Ziel der Akzeptanz bei den von der Evaluation Betroffenen führt.

In der vorliegenden Arbeit werden erste Schritte auf diesem Weg über eine Expertenbefragung unternommen. Die empirische Ausgangsfrage bezieht sich darauf, ob es für die Gruppe der an deutschen Psychologischen Instituten mit Hauptfachausbildung tätigen Professor(innen) und Privat-Dozenten(innen) gelingt, eine hinreichend konsensuelle Liste von Evaluationskriterien für die eigene berufliche Leistung (im universitären Kontext) zu identifizieren. Eine nachgeschaltete Zusatzfrage betrifft die empirische Sinnhaftigkeit der Dimensionierung und Aggregation der (konsensuellen) Evaluations- oder Produktivitätskriterien zu (psychometrischen) Skalen.

Den Ausgangspunkt der Liste möglicher Evaluationskriterien für die Tätigkeit von Hochschullehrern(innen) im Hauptfachstudiengang Psychologie bildeten in der Fachliteratur vorliegende Übersichten zu Indikatoren für die Leistungsmessung und -bewertung in der Forschung und Lehre (vgl. Alewell, 1988; Bolsenkötter, 1986; Daniel, 1988; 1996; Daniel & Fisch, 1988; Fisch & Daniel, 1986; House, 1993; Spada, 1997; van Raan, 1996a; Weingart & Winterhager, 1984; Weingart, Sehringer & Winterhager, 1991). Diese – häufig lediglich kursorisch oder taxonomisch konzeptualisierten, selten empirisch generierten oder überprüften – Listen von Evaluationskriterien wurden im *ersten Schritt* nach den Prinzipien (a) der Reduktion aufgrund hoher Ähnlichkeit bzw. Identität von Kriterien und (b) des Erhalts einer möglichst hohen inhaltlichen Vielfalt der Kriterien (d.h., Bandbreite) zusammengefaßt. Im *zweiten Schritt* wurde diese Liste diskursiv erweitert, wobei vor allem auf die oben zitierte Literatur zur Kritik an vorliegenden „ranking“-Studien sowie zu den Möglichkeiten und zur grundsätzlichen Kritik an den Möglichkeiten einer leistungsbezogenen Mittelvergabe im universitären Bereich (vgl. etwa Felt et al., 1995; Frühwald, 1997; Heckhausen, 1988; Süllwold, 1997b; van Raan, 1996b) Bezug genommen wurde. Der *dritte Schritt*, über dessen Ergebnisse im folgenden berichtet wird, bezieht sich auf eine Expertenbefragung, in deren Rahmen alle zu einem Stichtag an deutschen Psychologischen Institut mit Hauptfachausbildung tätigen Prof. und PD mit der Bitte angeschrieben wurden, eine aus 117 Einzelitems bestehende Liste möglicher Evaluationskriterien für die Tätigkeit an Hoch-

schulinstiuten persönlich nach ihrer jeweiligen Wichtigkeit zu bewerten. Um Aussagen über die Konsensfähigkeit der einzelnen Kriterien und die auf sie bezogenen sozialen Orientierungslagen der Befragten zu ermöglichen, wurde neben der persönlichen Bewertung jedes Kriteriums auch die bei der Mehrheit der deutschsprachigen Kollegen und Kolleginnen vermutete Bewertung jedes Evaluationskriteriums (d.h., die vermutete Gruppenbewertung) erfragt. Aus dem Vergleich dieser Urteilsvariablen (Selbsturteil und vermutetes Gruppenurteil) können nach der Konformitätstypenanalyse (KTA; Breznitz, 1967) anhand der Informationen über subjektive und objektive Bewertungskonformitäten und -diskonformitäten vier unterschiedliche soziale Orientierungslagen und drei Indizes zur Genauigkeit der Kollegenwahrnehmung abgeleitet werden.

Exkurs: Die Konformitätstypenanalyse (KTA)

Für die empirische Analyse sozialer Orientierungslagen (und damit sozialer Kognitionen) liegt mit der von Breznitz (1967) vorgestellten Konformitätstypenanalyse (KTA) ein Datenerhebungs- und Datenauswertungsformat vor, das eine der bedeutsamen bisherigen Beschränkungen des auf soziale Kognitionen gerichteten Forschungsprogramms überwinden helfen kann (vgl. hierzu vor allem auch Fisch, 1977; Krampen, 1994). Diese Beschränkung bezieht sich auf die weitgehend subjektivistische (oder mentalistische) Ausrichtung der Erforschung sozialer Kognitionen in der Sozial- und Entwicklungspsychologie. Als Elemente der „psychischen Realität“ von Individuen werden soziale Kognitionen bislang vor allem als subjektive Konstruktionen der sozialen Umwelt und der Beziehungen der eigenen Person zu ihr experimentell und anhand von Selbstauskünften analysiert. Externe Kriterien (zumal Validitätskriterien) fehlen, und die Forschung begnügt sich – wie zumeist im (sozial-)kognitivistischen Forschungsprogramm (siehe hierzu etwa auch Herrmann, 1982) – mit diesen Selbstauskünften, die allenfalls anhand anderer Selbstauskünfte (derselben Person) oder anhand der Effekte experimenteller Variationen validiert werden. Ausnahmen beziehen sich auf Validierungsversuche über den Vergleich von Selbst- und Fremdwahrnehmungsdaten (etwa bei Dyaden; vgl. hierzu etwa Krauss & Fussell, 1991), wobei Verzerrungen bei den Fremdwahrnehmungsdaten mindestens ebenso wahrscheinlich sind wie bei den Selbstauskünften (siehe hierzu auch Schneider, 1991). Nötig ist die Ergänzung dieses mentalistischen Forschungszugangs, in dem allzu oft rein pragmatisch und wenig reflektiert darauf vertraut wird, daß Menschen „good-enough perceivers“ (Fiske, 1993, p. 155) sind, um externe oder zumindest quasi-externe Variablen, die nicht von der jeweils untersuchten Person allein abhängen.

Breznitz (1967) hat nun eine einfach handhabbare, höchst transparente und zudem ökonomische Methodik vorgeschlagen, diese Problematik im Bereich sozialer Konstruktionen durch die Analyse der subjektiven wie

auch der objektiven Konformität von Individuen in Einstellungen, Werthaltungen, Vorurteilen etc. zu lösen. Nach Breznitz (1967) können bei Vorgabe alternativer Antwortmöglichkeiten (oder bei Dichotomierung anders skalierten Daten) aus der Beziehung zwischen der persönlichen Meinung eines Befragten zu einem bestimmten Urteilsobjekt und seiner Ansicht über die vorherrschende Mehrheitsmeinung zu dem betreffenden Objekt (statistische Gruppennorm; ggfs. bezogen auf eine definierte Bezugsgruppe wie etwa die Kollegen, die Kommilitonen o. ä.) vier verschiedene Antworttypen a priori abgeleitet werden. Diese Antworttypen können im Deutschen (in Anlehnung an Orlik, Fisch & Saterdag, 1971; Fisch, 1977) wie folgt beschrieben werden:

– *Typ I (wissentlich konform oder voll integriert)*: Die Mehrheitsmeinung ist richtig erfaßt, die persönliche Meinung und die statistische Gruppennorm stimmen objektiv und subjektiv überein. Fisch (1977) bezeichnet diese Orientierungslage als ideal für Interaktionen innerhalb von Gruppen, weil die Person ein zutreffendes Meinungsbild hat und keine Meinungsverschiedenheiten zwischen sich und der Gruppe wahrnimmt.

– *Typ II (vermeintlich konform)*: Die Mehrheitsmeinung ist falsch erfaßt, die persönliche Meinung und die vermutete Mehrheitsmeinung stimmen lediglich subjektiv, aber nicht objektiv überein. Fisch (1977) bewertet diese Orientierungslage als problematisch, da etwa dann Schwierigkeiten entstehen werden, wenn ein vermeintlich Konformer für die Gruppe handeln möchte.

– *Typ III (vermeintlich abweichend oder unwissentlich konform)*: Die Mehrheitsmeinung ist falsch erfaßt, die persönliche Meinung und die vermutete Mehrheitsmeinung weichen subjektiv voneinander ab, stimmen aber objektiv überein. Auch diese Orientierungslage der Selbstwahrnehmung als Außenseiter, ohne es aber zu sein, ist nach Fisch (1977) problematisch, da sie mit der Gefahr von Kräfteverlusten verbunden ist, die objektiv unnötig sind.

– *Typ IV (wissentlich abweichend)*: Die Mehrheitsmeinung ist richtig erfaßt, die persönliche Meinung und die vermutete Mehrheitsmeinung stimmen sowohl subjektiv als auch objektiv nicht überein. Diese Orientierungslage ist – „abgesehen von möglichen Problemen des Außenseitertums“ – „insofern nicht ungünstig, als Selbsteinschätzung und Beurteilung der anderen übereinstimmen“ (Fisch, 1977, S. 139).

Neben diesen vier sozialen Orientierungslagen ermöglicht die KTA die Ableitung von drei einfachen Indices über die Genauigkeit der Personen- bzw. Gruppenwahrnehmung:

– *Objektive Konformität*: Die Meinung einer Person stimmt mit der tatsächlichen Mehrheitsmeinung (statistische Norm) überein.

– *Subjektive Konformität*: Die Meinung einer Person stimmt mit der vermuteten Gruppenmeinung überein.

– *Kenntnis der (statistischen) Norm*: Die von einer Person vermutete Gruppenmeinung stimmt mit der tatsächlichen Mehrheitsmeinung (statistische Norm) überein.

Empirische Basis der KTA sind dichotome Daten über die persönliche Meinung und die vermutete Mehrheitsmeinung (in einer Bezugsgruppe) zu einem Einstellungs- oder Bewertungsobjekt (das sich natürlich auch auf Personen oder Gruppen beziehen kann). Liegen metrische Daten vor, so müssen Dichotomierungen vorgenommen werden, die anhand eines absoluten Kriteriums (z. B. Zusammenfassung aller (eher) zustimmenden Antwortkategorien und aller (eher) ablehnenden Antwortkategorien zu den dichotomen Kategorien der Bejahung versus Verneinung) durchzuführen sind. Die Verteilungen der (dichotomen bzw. dichotomisierten) Daten für das Selbsturteil und das vermutete Gruppenurteil (die vermutete Mehrheitsmeinung) können dann in Vier-Felder-Tafeln übertragen werden, aus denen sich direkt die Häufigkeitsverteilung der vier Konformitätstypen ablesen läßt. Der Modalwert der persönlichen Meinungen (bzw. Einstellungen, Bewertungen o. ä.) bezeichnet unter der Voraussetzung der Repräsentativität der Stichprobe (unter Bezug auf die – ggfs. enger – definierte Population) die statistische Gruppennorm. Die Zuordnung der Konformitätstypen ist dann von der Lokation des Modalwertes abhängig, wobei a priori zwei Fälle möglich sind, die in den Tabellen 1 und 2 demonstriert werden.

Eine Zuordnung der Konformitätstypen zu den Antworttypen (wie in den Tab. 1 und 2 demonstriert) ist nur dann statistisch abgesichert, wenn die Randsummen p und q hinreichend (d. h., statistisch bedeutsam bei $p < .01$ oder $p < .05$) differieren. Dies kann mit dem McNemar-Test (bivariater Zugang; siehe etwa Lienert, 1973, S. 191ff.) leicht geprüft werden.

Fassen wir zusammen: Als Ausgangsdaten für die KTA benötigt man dichotome (oder dichotomierte) Daten, die sich auf die persönliche Meinung (Einstellung, Bewertung o. ä.) und die vermutete Mehrheitsmeinung (in einer definierten Bezugsgruppe) zu einem Einstellungsobjekt (o. ä.) beziehen. Diese Daten werden in Kontingenztafeln der einfachsten Form (d. h., Vier-Felder-Tafeln) in Beziehung gesetzt, woraus sich bei Beachtung der statistischen Gruppennorm (auf der Basis der *wirklichen* Mehrheitsmeinung nach den Selbsturteilen) direkt die Verteilung der Befragten auf die vier Konformitätstypen und drei Konformitätsindices ergibt. Das Auswertungsverfahren und die benötigten nonparametrischen statistischen Testprozeduren sind höchst einfach zu handhaben (auch mit dem Taschenrechner) und zugleich höchst transparent sowie nachvollziehbar. Die KTA wurde bislang etwa erfolgreich in Analysen sozialer Orientierungslagen von Psychologiestudierenden (Orlik et al., 1971), Wissenschaftlern (Fisch, 1977) und Lehrern (Brandtstädter & Krampen, 1979; Krampen, Freilinger & Wilmes, 1994) sowie sozialer Orientierungslagen in den Bereichen der subjektiven Erklärungsmuster zur Humanentwicklung (Krampen, Freilinger & Wilmes, 1994) und zur Ausländerfeindlichkeit (Krampen & Krämer, 1995)

eingesetzt. Mit der Intergruppen-Wahrnehmungstypenanalyse (WTA; siehe Krampen, 1994) liegt zudem eine Erweiterung der KTA zu einer kombinierten intra- und interkulturell vergleichenden Analyseverfahren für soziale Kognitionen vor.

Methode

Untersuchungspopulation und Repräsentativität der Stichprobe

Als Untersuchungspopulation wurden alle Professor(innen) und Privat-Dozenten(innen) definiert, die nach dem „Psychologen-Kalender“ (Hogrefe-Verlag) des Jahres 1997 an deutschen Psychologischen Instituten mit Hauptfachausbildung („A-Institute“ nach dem Psychologen-Kalender) tätig waren. Die Populationsbeschränkung auf die „A-Institute“ der Bundesrepublik Deutschland und den genannten Personenkreis mit „*venia legendi*“ mußte aus ökonomischen Gründen vorgenommen werden. Im Juni 1997 wurde diese Gesamtpopulation von 555 Hochschullehrern(innen) im Hauptfach Psychologie mit der Bitte angeschrieben, gegen Honorierung einen umfangreichen Fragebogen zu bearbeiten, der neben Fragen zu möglichen Evaluationskriterien (siehe unten) Abschnitte mit differenzierten Fragen zur Nutzung wissenschaftlicher Informationsquellen und Literaturgattungen, zum Arbeits- und Publikationsverhalten, zu den Arbeitsbedingungen sowie zu persönlichen Daten (Alter, Geschlecht, Jahre der Promotion, Habilitation und Erstberufung sowie Anzahl weiterer angenommener und abgelehnter Berufungen) umfaßte. Für die Bearbeitung des gesamten Fragebogens war minimal eine Stunde anzusetzen.

Zwei Monate nach dem Fragebogenversand wurde ein Erinnerungsschreiben verschickt. Nach sechs Monaten hatten 265 Personen verwertbare Fragebogen zurückgeschickt, wobei die überwiegende Mehrheit (über 80 Prozent) innerhalb der ersten beiden Monate einging. Der Fragebogenrücklauf beträgt somit 48 Prozent und liegt

damit im Bereich vergleichbarer Expertenbefragungen bei Hochschullehrern (siehe etwa im Überblick Fisch, 1977, der für Hochschullehrer unterschiedlicher Fächer über Rücklaufquoten zwischen 44 und 89 Prozent berichtet). Für Stichproben von deutschsprachigen Hochschullehrern der Psychologie liegen Rücklaufquoten von 17 Prozent (Plaum, Schweitzer & Scharlach, 1989/90), 53 Prozent (Fisch, 1977) und 57 Prozent (Lonner, 1968) vor. Bei der hier realisierten Rücklaufquote sind zunächst der erhebliche Aufwand bei der Bearbeitung des sehr umfangreichen Fragebogens (23 Seiten Umfang) und die Tatsache, daß seine Rücksendung bei persönlicher Zusage von Vertraulichkeit, sofortiger Trennung von Namen und Fragebogen nach dem Eingang, Fragebogenvernichtung nach zwei Jahren sowie ausschließlicher Publikation nicht-namentlicher, über Personen aggregierter Ergebnisse nicht anonym erfolgte (dies schon deswegen, weil die zahlreichen individuellen Angaben ohnehin fallweise Identifikationen ermöglicht hätten und daher Vertrauen der Kolleginnen und Kollegen in die Projektgruppe Voraussetzung war), zu bedenken. Im Unterschied zu den anderen Untersuchungen ist ferner hervorzuheben, daß über Literaturdatenbanken und den „Psychologen-Kalender“ zu ausgewählten persönlichen und berufsbezogenen Variablen objektive Informationen über die gesamte Untersuchungspopulation vorliegen, die systematische Vergleiche der Fragebogen-Beantworter und Nicht-Antworter ermöglichen.

Objektive Daten zu Geschlecht, beruflichem Status (Prof. oder PD) und groben Indikatoren der Größe des Heimatinstituts (Anzahl der Professuren, Privat-Dozenturen, wissenschaftlichen Mitarbeiter(innen), Projektmitarbeiter(innen), Hauptfachstudierenden und Doktoranden(innen)) wurden dem Psychologen-Kalender (1997) entnommen. Aus den einschlägigen Literaturdatenbanken PSYINDEX (1977–1997) und PsycLit (1974–1997) wurden objektive Daten zum individuellen Publikationsaufkommen (Anzahl der nachgewiesenen Buchpublikationen, Buchbeiträge, Zeitschriftartikel und „grauen“ Publikationen) und zur Anzahl der betreuten Dissertationen ermittelt. Die individuelle Anzahl der Fremdzitationen und der Selbstzitationen von unter Erst-

Tabelle 1. KTA-Fall I der Zuordnung der Konformitätstypen zu den Antworttypen bei $p > q$ (d. h., die objektive Mehrheitsmeinung liegt bei der Zustimmung, p)

persönliche Meinung	vermutete Mehrheitsmeinung		Randsumme
	Zustimmung (ja)	Ablehnung (nein)	
Zustimmung (ja)	Typ I: wissentlich konform	Typ III: vermeintlich abweichend	p (= Modus bzw. Norm)
Ablehnung (nein)	Typ IV: wissentlich abweichend	Typ II: vermeintlich konform	q
Randsumme	p'	q'	N

Tabelle 2. KTA-Fall II der Zuordnung der Konformitätstypen zu den Antworttypen bei $p < q$ (d. h., die objektive Mehrheitsmeinung liegt bei der Zustimmung, q)

persönliche Meinung	vermutete Mehrheitsmeinung		Randsumme
	Zustimmung (ja)	Ablehnung (nein)	
Zustimmung (ja)	Typ II: vermeintlich konform	Typ IV: wissentlich abweichend	p
Ablehnung (nein)	Typ III: vermeintlich abweichend	Typ I: wissentlich konform	q (= Modus bzw. Norm)
Randsumme	p'	q'	N

autorenschaft publizierten Zeitschriftenbeiträgen wurden ebenso wie die Anzahl der zitierten Artikel nach dem SSCI (1973–1997) bestimmt. Bei allen bibliometrischen Datenbankanalysen wurden ggfs. vorliegende Varianten in der Schreibweise von Nachnamen und Vornamen (bzw. Initialien) soweit wie möglich berücksichtigt. In wenigen nicht zu klärenden Einzelfällen (vor allem bei Namensidentitäten) wurden die aus der jeweiligen Datenbank ermittelten Variablen auf fehlende Werte gesetzt.

Nach Tabelle 3 ist zu erkennen, daß sich die Fragebogen-Beantworter von den Nicht-Antwortern weder im beruflichen Status noch im Geschlecht statistisch bedeutsam unterscheiden. Auch für die aus dem SSCI ermittelten Zitationsindices liegen keine signifikanten Gruppenunterschiede vor (siehe Tabelle 4). Dies gilt ebenso nach Tabelle 4 für alle (groben) Indikatoren der Größe des Psychologischen Instituts, an dem die Antworter vs Nicht-Antworter tätig sind. Signifikante Unterschiede zwischen diesen Gruppen finden sich dagegen in einigen Indikatoren des Publikationsaufkommens (Tabelle 4). Während sich für die Anzahlen der im PSYINDEX dokumentierten Bücher, Buchbeiträge und „grauen“ Veröffentlichungen (Institutsreihen, Arbeitsberichte etc.) keine bedeutsamen Gruppenunterschiede zeigen, liegen statistisch signifikante Mittelwertsunterschiede für die PSYINDEX-Nachweise von Zeitschriftbeiträgen, für alle PsycLit-Nachweise und für die Anzahl der zitierten Zeitschriftbeiträge nach dem SSCI vor. Danach verfügen die Fragebogen-Beantworter im Vergleich zu den Nicht-Antwortern bei den zuletzt genannten objektiven Indikatoren des Publikationsaufkommens durchgängig über höhere Mittelwerte. Auch für die Anzahl der betreuten Dissertationen wird der Gruppenvergleich signifikant: Nicht-Antworter sind jedoch im PSYINDEX mit durchschnittlich vier Betreuungen bedeutsam häufiger dokumentiert als Fragebogen-Beantworter (mit durchschnittlich drei Betreuungen). Auffällig ist bei diesen fünf (von 17 analysierten) signifikanten Gruppenunterschieden auch, daß die erhöhten Mittelwerte dieser objektiven Produktivitätsindikatoren in vier Fällen (für die Zahl der zitierten Beiträge nach dem SSCI wird der Levene's F-Wert zur Varianzheterogenität bei $p < .06$ lediglich knapp nicht

signifikant) mit statistisch signifikant höheren Varianzen einhergehen. Erhöhte Gruppenmittelwerte in den objektiven Indikatoren treten also zusammen mit einer größeren Heterogenität innerhalb der jeweiligen Gruppe auf.

Zur Repräsentativität der Stichprobe für die untersuchte Population der Hochschullehrer(innen) an deutschen Psychologischen Instituten mit Hauptfachausbildung ist somit festzuhalten, daß sie für die Variablen Geschlecht, beruflicher Status, Größe des Heimatinstituts, Fremd- und Selbstzitationen nach dem SSCI sowie den Anzahlen der in PSYINDEX dokumentierten Buchveröffentlichungen, Buchbeiträgen und „grauen“ Veröffentlichungen empirisch abgesichert ist. Dies gilt jedoch nicht für die in PsycLit dokumentierte Anzahl von Veröffentlichungen, die Anzahl der zitierten Zeitschriftenbeiträge nach dem SSCI sowie für die in PSYINDEX dokumentierte Anzahlen der Zeitschriftenpublikationen und der erstbetreuten Dissertationen. Aufgrund der Repräsentativitätsmängel der Stichprobe in diesen objektiven Indikatoren wurde *innerhalb* der Stichprobe der Fragebogen-Beantworter geprüft, ob der jeweilige objektive Leistungsindikator statistisch bedeutsam mit der persönlichen Bewertung des auf diesen Indikator bezogenen Evaluationskriteriums korreliert ist.

Statistisch nicht signifikante Befunde zwischen der persönlichen Bewertung der Tauglichkeit als Evaluationskriterium und des objektiven aus Datenbanken gewonnenen Leistungskriteriums zeigen sich dabei für die PSYINDEX-Anzahl der Zeitschriftenbeiträge ($r = .10$) und der erstbetreuten Dissertationen ($r = .06$) sowie die PsycLit-Anzahl der Bücher und Buchbeiträge ($r = .00$). Damit ist zwar die Repräsentativität der Stichprobe nach diesen objektiven Indikatoren, in denen signifikante Unterschiede zu den Nicht-Antwortern bestehen, empirisch nicht abgesichert. Innerhalb der Stichprobe der Antworter können aber Verzerrungen bei den persönlichen Bewertungen der Evaluationskriterien aufgrund erhöhter Publikationszahlen und reduzierter Erstbetreuungen für Dissertationen ausgeschlossen werden. Anders ist dies bei den persönlichen Bewertungen und Indikatoren der PsycLit-Nachweise von Zeitschriftenbeiträgen ($r =$

Tabelle 3. Vergleiche der Fragebogen-Beantworter ($N = 265$) mit den Nicht-Antworten ($N = 290$) nach dem beruflichen Status und dem Geschlecht

Variable	Antworter		Nicht-Antworter		Chi ² (df = 1)
	f	%	f	%	
<i>beruflicher Status</i>					
Professor	218	82	241	83	0.07
Privat-Dozent	47	18	49	17	
<i>Geschlecht</i>					
weiblich	42	16	41	14	0.32
männlich	223	84	249	86	

.20, $p < .01$) und der SSCI-Anzahl zitierter Zeitschriftenbeiträge ($r = .20$, $p < .01$). Für diese beiden Variablen deuten sich somit innerhalb der Stichprobe schwache, jedoch signifikante Zusammenhänge zwischen den persönlichen Bewertungen der beiden darauf bezogenen Evaluationskriterien und den entsprechenden persönlichen Datenbanknachweisen von Zeitschriftenbeiträgen (in PsycLit und SSCI) an.

Fragebogen

Ein Abschnitt des umfangreichen Fragebogens bezog sich auf die Bewertung von 117 möglichen Evaluationskriterien für die Tätigkeit von Hochschullehrern(innen) der Psychologie (siehe Tabelle 5; zur Generierung dieser Liste s.o.). Die Befragten wurden gebeten, die einzelnen Kriterien auf neunstufigen Skalen (von „nicht positiv“ = 0 bis „höchst positiv“ = 8) (1) nach der persönlichen Wichtigkeit (Selbsturteil) und (2) nach der bei der Mehrheit ihrer deutschsprachigen Kollegen und Kolleginnen vermuteten Bewertung (vermutetes Gruppenurteil) einzustufen. Die Vorgabe bipolarer Antwortskalen wurde verworfen, da die Mehrzahl der Evaluationskriterien kaum (sehr) negativ bewertet werden dürfte (dafür steht summarisch beim verwendeten Format der Skalenpunkt 0 zur Verfügung) und für asymmetrische bipolare Antwortskalen keine hinreichenden Forschungserfahrungen vorliegen. Zur besseren Übersicht wurde die Kriterienliste grob inhaltlich nach den Themenbereichen (a) Veröffentlichungen, (b) Forschungsaktivitäten, (c) Berufungen und Ehrungen, (d) Lehre, (e) akademische Selbstverwaltung, (f) Gutachtertätigkeit und (g) Funktionen in Fachgesellschaften strukturiert, wobei explizit darauf verwiesen wurde, daß weder die Abfolge der vorgegebenen Evaluationskriterien noch deren Gliederung in einem Zusammenhang mit ihrer Gewichtung stehen. Zudem wurde zweimal innerhalb der Liste (bei den Aspekten der Qualitätsindikatoren für wissenschaftliche Arbeit und der Funktionen in Fachgesellschaften) sowie an ihrem Ende offen nach Ergänzungen der Auflistung von Evaluationskriterien gefragt, die persönlich vermißt werden.

Ergebnisse

Persönliche Bewertungen der Evaluationskriterien

Schon die grobe Inspektion der Mittelwerte der persönlichen Bewertungen der 117 vorgegebenen Evaluationskriterien in Tabelle 5 zeigt, daß ihre große Mehrheit im Selbsturteil der Befragten positiv beurteilt wird. Lediglich bei 16 Evaluationskriterien sind durchschnittlich *weniger positive Bewertungen* zu konstatieren, d. h., der Mittelwert liegt unter dem mittleren Skalenpunkt von 4: In aufsteigender Reihenfolge gilt dies für eine durchschnittlich niedrigere persönliche Bewertung (als Evaluationskriterium) von Newsletter-Publikationen, Publikationen „grauer“ Literatur, Softwarebesprechungen, Buchbesprechungen, Kongreß-/Tagungs-/Workshopberichten, Funktionen im BDP, Editorials, Newsletter-Editionen, Publikationen in Kongreß- und Tagungsbänden, Mitgliedschaften im Fachbereichs-/Fakultätsrat, Mitarbeit in anderen Gremien der akademischen Selbstverwaltung, der Anzahl der Rezensionen eigener Bücher, Publikationen von Schulungs-/Trainingsmaterialien, Erwähnungen in den Massenmedien, Posterpräsentationen auf nationalen Kongressen/Tagungen sowie Begutachtungen von Beitragsmeldungen für Kongresse.

Mittlere, insgesamt positive Bewertungen (um die Skalenwerte 4–5; siehe Tabelle 5) erhalten Evaluationskriterien, die sich etwa auf das Renommee zitierender Autoren, die Alleinautorenschaft, gemeinsame Publikationen mit renommierten Kollegen, die Publikation von „overviews“/Testmanualen/Multimedia Produkten/Softwareprodukten/“professional criticisms“/Repliken, Editionen von Sammelwerken und Buchreihen, das Alter bei Promotion/Habilitation/erster Berufung, die Anzahl der betreuten Diplomarbeiten und Dissertationen, Drittmitteleinwerbungen von Bund/Land/EU/Stiftungen/Wirtschaft, angemeldete Vorträge auf nationalen Kongressen/Tagungen/Workshops, die Organisation von Arbeitsgruppen auf nationalen/internationalen Kongressen/Tagungen, Posterpräsentationen auf internationalen Kongressen

Tabelle 4. Vergleich der Fragebogen-Beantworter ($N = 265$) mit den Nicht-Antwortern ($N = 290$) nach objektiven Indikatoren des Publikationsaufkommens, der Anzahl der erstbetreuten Dissertationen sowie der Größe des Heimat-instituts

Variable	Antworter		Nicht-Antworter		Levene's Test: F	df	t-Wert
	M	SD	M	SD			
<i>Veröff. (PSYNDEX)</i>							
Bücher	3.6	4.7	3.3	3.8	0.60	553	0.70
Buchbeiträge	9.5	9.6	8.2	8.9	0.45	553	1.66
Zeitschriftenbeiträge	17.4	15.9	12.7	11.6	8.09**	479	3.90**
„Graue“ Literatur	4.5	9.1	3.5	8.9	4.15*	546	1.25
<i>Veröff. (PsyncLIT)</i>							
Bücher/Buchbeiträge	1.9	2.9	1.4	2.4	4.54*	518	2.16*
Zeitschriftenbeiträge	14.5	13.8	9.5	10.8	11.68**	499	4.70**
<i>Zitationen</i>							
SSCI: Selbstzit.	101.4	108.1	93.1	171.9	0.39	532	0.66
SSCI: Fremdzit.	93.3	108.8	86.3	165.6	0.19	532	0.58
SSCI: zit. Beiträge	20.3	17.7	14.3	15.9	3.50	532	4.16**
<i>Betreute akad. Arbeiten</i>							
PSYNDEX:							
Dissertationen	3.2	5.2	4.3	6.7	10.69**	541	2.11*
<i>Heimatinstitut</i>							
Anz. der Professoren	11.8	4.6	12.4	4.7	0.94	525	1.48
Anz. der Privatdozenten	2.8	2.1	2.9	2.1	0.17	525	0.88
Anz. der wiss. Mitarbeiter	24.6	12.6	25.7	13.0	0.14	525	0.91
Anz. der Lehrbeauftragten	8.4	8.3	8.5	8.0	0.42	525	0.03
Anz. der Projektmitarbeiter	11.3	10.5	11.4	10.7	0.01	525	0.11
Anz. der Hauptfach- studierenden	689.6	281.3	691.4	311.9	0.23	339	0.05
Anz. der Doktoranden	27.2	10.9	28.0	11.9	2.08	330	0.62

Anmerkung: * $p < .05$, ** $p < .01$

sen/Tagungen, die Anzahl der wissenschaftlichen Mitarbeiter, die Anzahl der nationalen Kooperationen, die Berufung in wissenschaftliche Beiräte/Expertengremien, Honorarprofessuren, Stipendien, Gastprofessuren im Inland, Lehrstuhlvertretungen, Forschungsaufenthalte im Ausland, das Amt des Sprechers/Direktors von Instituten, das Amt des Dekans, Gutachtertätigkeiten für Forschungsprojekte/Buchmanuskripte/Stipendiumsanhträge sowie Funktionen in der DGPs und in internationalen wissenschaftlichen Gesellschaften beziehen (= 42).

In den Selbsturteilen liegen die Mittelwerte für die Bewertung aller anderen 59 zur Einschätzung vorgegebenen Evaluationskriterien im *sehr positiven Bereich* der Antwortskala. Dies gilt insbesondere für die Spitzenreiter (mit $7.1 > M > 6.3$; in abnehmender Reihenfolge) der eingeladenen Hauptvorträge auf internationalen Kongressen/Tagungen, der Originalität wissenschaftlicher Frage-

stellungen, der Zeitschriftenpublikationen, des Renommées der Zeitschriften, der Gründung eines Forschungsschwerpunkts, des Betreibens eines Forschungszentrums, der didaktischen Aufbereitung und Qualität von Lehrveranstaltungen, des Engagements/Commitments in der Lehre, der Aktualität von Lehrveranstaltungen, der individuellen Betreuung und Beratung von Studierenden, der von der DFG eingeworbenen Drittmittel sowie der Publikation empirischer und experimenteller Untersuchungen (siehe Tabelle 5).

Unter Bezug auf die Diskussion um die internationale Dissemination psychologischer Befunde und Theorien aus dem deutschsprachigen Bereich (vgl etwa Keul et al., 1993, 1994; Montada et al., 1995) sei darauf verwiesen, daß bei allen Evaluationskriterien, die separat nach nationalen versus internationalen Aktivitäten (wie Kongreß- und Tagungsbesuche, wissenschaftliche Kooperationen,

Tabelle 5. Deskriptive Parameter der Evaluationskriterien im Selbsturteil und im vermuteten Gruppenurteil sowie prozentuale Verteilung der Konformitätstypen und -indices

Evaluationskriterium	Selbsturteil		vermutetes Gruppenurteil		Norm ^a	Konformitätstyp (%) ^b				Konformitätsindex (%) ^c		
	M	SD	M	SD		I	II	III	IV	OK	SK	KN
Anzahl d. Publikationen	5.5	1.7	6.0	1.5	****	84	2	7	8	91	86	92
Publikationen/Berufsdauer	6.1	1.6	5.9	1.5	****	88	1	6	6	94	89	94
<i>Wirkung der Publikationen</i>												
Anz. der (Fremd-)Zitationen	5.7	1.8	5.8	1.6	****	81	2	7	10	88	83	91
Anz. der Cited Core Documents	5.2	1.9	5.6	1.7	****	73	7	7	14	80	80	87
Anz. der Buchrezensionen	3.6	2.1	4.2	1.9	**	44	43	11	2	55	87	46
Anz. der Erwähnungen i. syst. Reviews	5.9	1.7	5.8	1.6	****	86	5	5	4	91	91	90
Anz. der Erwähnungen i. Lehrbüchern	5.8	1.7	5.9	1.5	****	89	5	3	3	92	94	92
Anz. der Erwähnungen i. Massenmedien	3.9	2.1	4.3	2.0								
Renommee d. zitierenden Autoren	4.5	1.9	5.4	1.7	****	71	16	1	12	72	87	83
Rang d. zitierenden Publikationsquellen	5.7	1.9	5.8	1.7	****	80	5	8	7	88	85	87
<i>Status des Autors</i>												
Alleinautorenschaft	4.8	2.1	5.5	1.6	****	70	8	3	19	73	78	89
Erstautorenschaft	5.2	2.0	5.8	1.4	****	79	6	2	13	81	85	92
Coautor renommierter Kollegen	4.8	1.9	5.7	1.4	****	75	8	2	15	77	83	90
<i>Publikationsort</i>												
Zeitschrift	6.9	1.2	6.5	1.4	****	97	1	3	0	100	98	97
Monographie	6.2	1.4	6.3	1.3	****	96	1	1	3	97	97	99
Sammelwerk	5.0	1.4	5.2	1.4	****	80	7	6	7	86	87	87
Lehrbuch	5.9	1.6	6.0	1.5	****	89	2	3	5	92	91	94
Enzyklopädie	5.6	1.7	6.1	1.4	****	88	2	1	9	89	90	97
Kongreß- oder Tagungsband	3.2	1.8	4.0	1.7	**	54	30	13	3	67	84	57
„Graue Literatur“	2.4	1.9	2.6	1.6								
Newsletter	2.2	1.8	2.7	1.7								
<i>Rang des Publikationsortes</i>												
Renommee der Zeitschrift	6.8	1.4	6.8	1.3	****	94	0	3	3	97	94	97
Renommee d. Hrsg. eines Sammelbandes	5.0	1.8	5.8	1.5	****	81	8	1	11	82	89	92
Renommee des Verlages	4.0	2.1	5.0	1.8	****	57	24	2	17	59	81	74

Fortsetzung von Tab. 5 auf der nächsten Seite

Tabelle 5. Fortsetzung

Evaluationskriterium	Selbsturteil		vermutetes Gruppenurteil		Norm ^a	Konformitätstyp (%) ^b				Konformitätsindex (%) ^c		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		I	II	III	IV	<i>OK</i>	<i>SK</i>	<i>KN</i>
<i>Kategorie der Publikation</i>												
Empirische Untersuchung	6.4	1.4	6.3	1.2	****	95	1	3	1	98	96	96
Experimentelle Untersuchung	6.4	1.7	6.5	1.3	****	91	0	3	7	94	91	98
Metaanalyse	5.5	1.7	5.7	1.4	****	84	5	3	8	87	89	92
Systematisches Review	6.0	1.3	5.8	1.3	****	93	2	3	2	96	95	95
Overview	5.2	1.4	5.2	1.4	****	85	7	6	3	91	92	88
Testentwicklung, Test-Handanweisung	5.0	1.9	5.3	1.4	****	76	8	5	11	81	84	87
Theoretische Analyse	6.1	1.5	5.6	1.5	****	86	2	8	5	94	88	81
Multimediale Produkte	4.0	1.8	4.0	1.6								
Softwareprodukte	4.0	1.9	4.0	1.7								
Buchbesprechung	2.7	1.7	3.3	1.7								
Softwarebesprechung	2.5	1.7	2.9	1.7	-*	78	13	7	2	85	91	80
Editorials	3.2	1.8	3.7	1.8	-***	53	31	14	2	67	84	55
Diskussionsbeiträge, Repliken	4.2	1.7	4.3	1.7	***	52	30	10	8	62	82	60
Schulungs- u. Trainingsverfahren/-mat.	3.7	1.9	3.6	1.7								
Berichte über Kongresse	2.7	1.8	3.2	1.8								
<i>Herausgeberschaft</i>												
Herausgeberschaft von Sammelbänden	4.9	1.6	5.4	1.4	****	77	7	3	12	80	84	89
Herausgeberschaft von Buchreihen	4.7	1.8	5.4	1.5	****	70	13	3	14	73	83	84
Herausgeberschaft von Zeitschriften	6.1	1.5	6.3	1.4	****	93	1	1	4	94	94	97
Herausgeberschaft von Newsletter	3.2	1.9	3.7	1.8	-*	58	32	8	2	66	90	60
<i>Qualität der wissenschaftlichen Arbeiten</i>												
Originalität der Fragestellung	7.3	1.0	6.3	1.5	****	93	0	6	0	99	93	93
Benutzung moderner Analysemethoden	5.8	1.7	6.2	1.4	****	89	3	1	7	90	92	96
Sorgfalt der Stichprobengewinnung	6.3	1.4	5.9	1.5	****	90	2	5	3	95	92	93
Aufwand der empirischen Untersuchung	6.3	1.5	6.1	1.4	****	90	0	4	5	94	90	95
Kreativität der Interpretation	6.3	1.4	5.5	1.5	****	87	3	7	4	94	90	91

Tabelle 5. Fortsetzung

Evaluationskriterium	Selbsturteil		vermutetes Gruppenurteil		Norm ^a	Konformitätstyp (%) ^b				Konformitätsindex (%) ^c		
	M	SD	M	SD		I	II	III	IV	OK	SK	KN
Alter bei Abschluß der Promotion	4.8	2.0	5.4	1.4	+***	74	7	3	15	77	81	89
Alter bei Abschluß der Habilitation	4.9	2.0	5.6	1.4	+***	75	7	3	15	78	82	90
Alter bei erster Berufung	4.7	2.0	5.5	1.4	+***	74	8	2	16	76	82	90
Anz. der betreuten Diplomarbeiten	4.4	1.9	4.3	1.7	+***	50	23	15	12	65	73	62
Anz. der betreuten Promotionen	5.2	1.7	5.3	1.5	+***	78	8	7	8	85	86	86
Anz. der betreuten Habilitationen	5.5	1.7	5.7	1.6	+***	85	4	3	7	88	89	92
<i>Höhe der eingeworbenen Drittmittel</i>												
Von Bundesministerien	5.5	1.8	6.1	1.3	+***	84	2	3	11	87	86	95
Von Landesministerien	5.2	1.8	5.9	1.3	+***	79	3	3	15	82	82	94
Von der EU	5.4	1.9	6.1	1.4	+***	82	2	3	13	85	84	95
Von der DFG	6.4	1.9	7.0	1.1	+***	91	0	2	7	93	91	98
Von Stiftungen	5.1	1.8	5.5	1.4	+***	82	6	3	9	85	88	91
Aus der Wirtschaft	4.9	2.0	5.3	1.5	+***	75	7	6	11	81	82	86
Aus sonstigen Quellen	4.6	1.8	5.0	1.5	+***	72	2	2	9	74	74	81
Insgesamt	6.0	1.9	6.5	1.3	+***	90	3	1	6	91	93	96
<i>Teilnahme an nationalen wiss. Kongressen</i>												
Angemeldeter Hauptvortrag	5.2	1.7	5.8	1.3	+***	86	5	1	9	87	91	95
Eingeladener Hauptvortrag	6.2	1.7	6.8	1.1	+***	93	2	0	5	93	95	98
Überblicks- oder Positionsreferat	5.4	1.6	5.8	1.2	+***	91	4	0	5	91	95	96
Forschungsreferat	5.1	1.6	5.2	1.4	+***	84	9	2	5	86	93	89
Organisation einer Arbeitsgruppe	4.7	1.7	5.2	1.4	+***	75	11	4	10	79	86	85
Poster	3.9	1.8	3.5	1.6	-**	44	32	6	18	50	76	62
<i>Teilnahme an internat. wiss. Kongressen</i>												
Angemeldeter Hauptvortrag	6.0	1.6	6.3	1.3	+***	93	1	1	4	94	94	97
Eingeladener Hauptvortrag	7.0	1.5	7.2	1.2	+***	96	0	0	3	96	96	99
Überblicks- oder Positionsreferat	6.0	1.3	6.3	1.2	+***	94	1	1	4	95	95	98
Forschungsreferat	5.7	1.4	5.8	1.3	+***	92	2	2	3	94	94	95
Organisation einer Arbeitsgruppe	5.4	1.6	5.6	1.4	+***	84	5	3	7	87	89	91
Poster	4.4	1.9	4.0	1.7	+**	48	30	15	7	63	78	55

Fortsetzung von Tab. 5 auf der nächsten Seite

Tabelle 5. Fortsetzung

Evaluationskriterium	Selbsturteil		vermutetes Gruppenurteil		Norm ^a	Konformitätstyp (%) ^b				Konformitätsindex (%) ^c		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		I	II	III	IV	<i>OK</i>	<i>SK</i>	<i>KN</i>
Anz. der Forschungsprojekte/Lebensalter	5.8	1.9	6.0	1.4	+***	87	4	1	8	88	91	96
Anz. der forschungsaktiven Mitarbeiter	5.1	2.1	5.7	1.6	+***	79	11	1	9	80	90	88
<i>Anzahl gemeinsamer Forschungsprojekte</i>												
Anz. nationaler Kooperationen	5.1	1.7	5.6	1.4	+***	84	6	4	6	88	92	92
Anz. internationaler Kooperationen	6.0	1.6	6.3	1.4	+***	92	1	2	5	94	93	97
Gründung eines Forschungsschwerpunktes	6.6	1.4	6.8	1.3	+***	97	1	1	1	98	98	98
Betreiben eines Forschungszentrums	6.6	1.5	6.9	1.3	+***	95	2	1	2	96	97	97
Zitierung in anderen Disziplinen	5.7	1.8	5.4	1.9	+***	76	11	11	2	87	87	78
<i>Berufungen und Ehrungen</i>												
Anz. der Berufungen auf Professuren	5.4	1.9	6.7	1.3	+***	83	1	1	15	84	84	98
Berufung in Akademien	5.6	1.9	6.5	1.5	+***	84	4	1	11	85	88	95
Berufung in wissenschaftliche Beiräte	5.1	1.7	6.1	1.4	+***	83	4	2	11	85	87	94
Berufung in wiss. Expertengremien d. Politik	4.8	1.8	5.8	1.5	+***	75	6	2	18	77	81	93
Berufung in Gremien von Wissenschaftsorg.	5.3	1.7	6.2	1.4	+***	86	2	1	10	87	88	96
Berufung in Gremien berufsständischer Org.	4.0	1.8	4.9	1.6	+***	51	18	4	27	55	69	88
Rang der berufenden Institutionen	5.6	1.7	6.3	1.4	+***	89	3	1	7	90	82	96
Wissenschaftliche Preise	5.9	1.7	6.7	1.2	+***	90	1	1	8	91	91	98
Ehrendoktorwürde	5.4	2.3	6.5	1.6	+***	80	6	1	14	81	86	94
Honorarprofessur	4.5	2.1	5.5	1.8	+***	67	14	1	18	68	91	95
Stipendien	5.2	1.7	5.5	1.5	+***	82	10	3	5	85	92	87
<i>Gastprofessuren</i>												
Gastprofessuren im Inland	5.1	1.7	5.7	1.5	+***	84	9	1	6	85	93	90
Gastprofessuren im Ausland	6.1	1.7	6.5	1.4	+***	91	3	2	4	93	94	95
<i>Lehrstuhlvertretungen</i>												
Anz. d. Forschungsaufenthalte im Ausland	5.2	1.9	5.6	1.5	+***	80	6	3	12	83	86	92
Anz. fremdfinanzierter Auslandsaufenthalte	5.7	1.7	6.0	1.4	+***	88	1	2	9	90	89	97
Reputationseinstufung durch Fachkollegen	5.6	1.8	6.4	1.3	+***	89	3	1	7	90	82	96

Tabelle 5. Fortsetzung

Evaluationskriterium	Selbsturteil		vermutetes Gruppenurteil		Norm ^a	Konformitätstyp (%) ^b				Konformitätsindex (%) ^c		
	M	SD	M	SD		I	II	III	IV	OK	SK	KN
<i>Lehre</i>												
Engagement/Commitment in der Lehre	6.5	1.4	4.8	1.7	+***	72	3	23	2	95	75	74
Breite der Lehrerfahrung	6.2	1.3	4.7	1.7	+***	73	3	23	2	96	76	75
Didaktische Aufbereitung u. Qualität	6.6	1.3	4.7	1.8	+***	72	2	26	1	98	74	73
Individuelle Betreuung der Studierenden	6.4	1.4	4.5	1.8	+***	68	3	28	2	96	71	70
Praxis-/Anwendungsorientierung	5.5	2.0	4.3	1.8	+***	53	10	28	9	91	63	62
Aktualität der Lehrveranstaltungen	6.4	1.5	5.1	1.8	+***	79	3	17	2	96	82	81
<i>Ämter in der akademischen Selbstverwaltung</i>												
Mitglied im Fachbereichsrat/Fakultätsrat	3.4	2.0	4.0	1.7	-*	46	34	14	6	60	81	52
Fachsprecher/Institutsdirektor	4.0	2.0	4.7	1.7	+***	54	25	3	18	57	79	72
Dekan und vergleichbare Positionen	4.3	2.1	5.1	1.7	+***	59	21	2	19	61	80	78
Mitarbeit in anderen Gremien	3.5	2.0	3.9	1.8	-*	43	35	14	7	57	78	50
<i>Häufigkeit der Gutachtertätigkeiten</i>												
Gutachtertätigkeit für Projektanträge	5.3	1.6	5.7	1.4	+***	85	4	3	8	88	89	93
Gutachtertätigkeit für Zeitschriftenmanuskripte	5.5	1.6	5.6	1.5	+***	86	7	2	5	88	93	91
Gutachtertätigkeit für Buchmanuskripte	4.7	1.6	5.1	1.5	+***	75	13	4	8	79	88	83
Gutachtertätigkeit für Stipendiansanträge	4.3	1.7	4.5	1.7	+***	61	32	2	6	63	93	67
Gutachtertätigkeit für Kongreßbeiträge	3.9	1.8	4.4	1.8	+**	55	36	2	8	57	91	63
Tätigkeit als DFG-Fachgutachter	6.1	1.7	6.9	1.2	+***	93	0	0	6	93	93	99
<i>Wichtige Funktionen in Fachgesellschaften</i>												
DGPs	4.3	2.1	5.9	1.5	+***	65	8	0	27	65	73	92
BDP	3.1	1.9	4.1	1.9	-***	42	28	26	4	68	70	46
Intern. wissenschaftliche Gesellschaften	5.2	1.9	5.9	1.5	+***	83	7	1	9	84	90	92

Anmerkungen: ^a + = sehr positive Bewertung oder - = weniger positive Bewertung in der Mehrheit der Selbsturteile mit Signifikanz bei * p < .01, ** p < .01, *** p < .001; ^b Konformitätstypen: I = voll integriert; II = vermeintlich konform; III = unwissentlich konform; IV = wissentlich nicht konform; ^c Konformitätsindices: OK = objektive Konformität; SK = subjektive Konformität; KN = Kenntnis der Norm.

Gastprofessuren sowie Funktion in Fachgesellschaften) zu bewerten waren, die internationalen Aktivitäten durchgängig höhere Mittelwerte erzielen. Die Differenzen in den persönlichen Einschätzungen liegt für die meisten dieser Evaluationskriterien über dem Wert einer halben Standardabweichung zu Gunsten einer positiveren Bewertung der internationalen Aktivitäten.

Zusammenfassend kann somit zunächst festgehalten werden, daß die deutliche Mehrheit der 117 vorgegebenen Evaluationskriterien im Durchschnitt als „positiv“ (36 Prozent) bzw. „höchst positiv“ (50 Prozent) bewertet wird, lediglich 14 Prozent der vorgegebenen Kriterien wird im Mittel als „wenig positiv“ beurteilt. Dies bestätigt die in der Einleitung aufgeführten grundsätzlichen Bedenken gegenüber singulären Evaluationskriterien und gegenüber kurzen Kriterienlisten empirisch. Die ebenfalls in Tabelle 5 aufgeführten Standardabweichungen weisen jedoch auf eine markante Urteilsvarianz bei der Mehrzahl der Evaluationskriterien, der im Zusammenhang mit den auch in Tabelle 5 aufgeführten Mittelwerten und Standardabweichungen der vermuteten Kollegenbewertungen im Rahmen der Konformitätstypenanalyse unter Konzentration auf die Frage nach der Konsensfähigkeit der Kriterien im nächsten Abschnitt der Befunddarstellung differenziert nachgegangen wird.

Freie Ergänzungen zu den zur Bewertung vorgegebenen Evaluationskriterien wurden zum Bereich der Qualität wissenschaftlicher Arbeiten von 33 Prozent und zum Bereich der Funktionen in Fachgesellschaften von 5 Prozent der Stichprobe vorgenommen. Insgesamt 28 Prozent der Stichprobe legte ferner weitere Ergänzungsvorschläge vor, nach denen am Ende der vorgegebenen Liste offen gefragt worden war. Während sich die wenigen Ergänzungen zu den Funktionen in Fachgesellschaften ausschließlich auf die Angabe von Funktionen in anderen als den im Fragebogen explizit benannten wissenschaftlichen Vereinigungen beziehen, zeigt die induktiv realisierte Inhaltsanalyse der anderen inhaltlichen Ergänzungen, daß vor allem Evaluationskriterien wie „die theoretische Einbettung/das theoriegeleitete Vorgehen“ ($n = 33$), „der Einbezug anderer Fächer/die Interdisziplinarität“ ($n = 27$), „die praktische und/oder gesellschaftliche Relevanz/der Anwendungsbezug“ ($n = 27$), und „die Lesbarkeit/Verständlichkeit“ ($n = 11$) wissenschaftlicher Arbeiten häufig frei benannt wurden. Geringere Nennhäufigkeiten ($n < 9$) finden sich für insgesamt 47 weitere Ergänzungsvorschläge, die inhaltlich sehr breit gestreut sind und von der Teamfähigkeit über Führungsqualitäten, die Kontinuität der Forschungstätigkeit, die Orientierung am „mainstream“ der Forschung und die Lehrevaluation durch Hörer bis zur Residenzpflicht reichen (die Liste dieser Ergänzungsvorschläge wird auf Anfrage gerne zur Verfügung gestellt).

Soziale Orientierungslagen bei der Bewertung der Evaluationskriterien

Da die Bestimmung der Konformitätstypen und -indices nach der KTA alternative Antwortkategorien voraussetzt,

wurden die Antwortverteilungen der Selbsturteile und vermuteten Gruppenurteile dichotomiert: Skalenwerte von 5–8 wurden der Kategorie „sehr positiv“ (+), Skalenwerte von 0–3 der Kategorie „weniger positiv“ (–) zugeordnet und Skalenwerte von 4 (mittlerer Skalenpunkt) wurden wegen ihrer Indifferenz fallweise aus der jeweiligen KTA ausgeschlossen. Die anhand der aus den so dichotomierten Selbst- und vermuteten Gruppenurteilen resultierenden Kontingenztafeln für alle 117 zur Bewertung vorgegebenen Evaluationskriterien berechneten KTAs führten zu den in Tabelle 5 vollständig dokumentierten Befunden.

Die Ergebnisse in Tabelle 5 zeigen zunächst, daß lediglich für acht der 117 Evaluationskriterien aufgrund der Selbsturteile nach dem McNemar-Test (bivariater Zugang; siehe Lienert, 1973) keine statistische (soziale) Norm inferenzstatistisch abgesichert werden konnte. D. h., daß für diese Kriterien (Erwähnungen in den Medien, Publikation „grauer“ Literatur und in Newslettern sowie Veröffentlichung von Multimedia Produkten, Softwareprodukten, Buchbesprechungen, Schulungs-/ Trainingsmaterialien und Kongreß-/Tagungsberichten) in der befragten Stichprobe keine eindeutige Mehrheitsmeinungen (im Sinne sozialer Normen) vorliegen. Die Berechnung von Konformitätstypen und -indices erübrigt sich daher für diese Items.

Für alle anderen 109 Evaluationskriterien wird die statistische Norm dagegen (zumeist hoch) signifikant. Bei 100 der Kriterien entspricht die soziale Norm einer sehr positiven Bewertung (+), und die Zuweisung der Konformitätstypen muß nach dem in Tabelle 1 demonstrierten Prinzip erfolgen. Bei neun Evaluationskriterien entspricht die soziale Norm dagegen einer weniger positiven Bewertung (–), und die Typenzuweisung hat dem in Tabelle 2 demonstrierten Prinzip zu folgen. Konsensuell weniger positiv bewertet werden demnach die Evaluationskriterien der Anzahl von Rezensionen eigener Bücher, der Publikation von Beiträgen in Kongreß- oder Tagungsbänden, der Software-Besprechungen, der Editorials, der Herausgabe von Newslettern, der Präsentation von Postern auf nationalen wissenschaftlichen Kongressen oder Tagungen, der Mitgliedschaft im Fachbereichs- oder Fakultätsrat, der Mitarbeit in anderen universitären Gremien der Selbstverwaltung sowie der Funktionen im Berufsverband Deutscher Psychologinnen und Psychologen (BDP).

Zu konstatieren bleibt zunächst, daß für die deutliche Mehrheit der zur Bewertung vorgegebenen Evaluationskriterien (nämlich 93 %) eine ausgeprägte Mehrheitsmeinung (soziale Norm) inferenzstatistisch abgesichert ist. Hier liegt somit unter den Befragten ein breiter Konsens in der sehr positiven versus weniger positiven Bewertung der Kriterien für die Beurteilung der Tätigkeiten von Hochschullehrern(innen) der Psychologie vor. Die Inspektion der Häufigkeitsverteilungen der vier Konformitätstypen und der drei Konformitätsindices (siehe Tabelle 5) verstärkt und differenziert dieses Befundmuster, da in der überwiegenden Mehrheit der Evaluationskriterien der wissentlich konforme/voll integrierte

Konformitätstyp I und die hohe Kenntnis der Norm die höchsten relativen Häufigkeiten aufweisen. Im Durchschnitt gehören 78 Prozent der Stichprobe dem sozial voll integrierten Typus an. D.h., Selbsterteil und Gruppenurteil stimmen subjektiv und objektiv überein, drei Viertel der Stichprobe hat somit (im Mittel über alle 109 Kriterien) ein zutreffendes Meinungsbild und nimmt keine Meinungsunterschiede zwischen sich und der Kollegenschaft wahr. Da auch der wissentlich abweichende Konformitätstyp IV die soziale Norm richtig erfaßt, liegt der mittlere Prozentsatz für die Kenntnis der Norm mit 87 Prozent der Stichprobe noch höher. Zugleich wird deutlich, daß (gemittelt über die 109 Evaluationskriterien) lediglich neun Prozent der Stichprobe dem Konformitätstypus IV angehört, mit dem bewußte Außenseiter beschrieben werden, die die Mehrheitsmeinung richtig erfassen, in der persönlichen Meinung aber von dieser Mehrheitsmeinung sowohl subjektiv als auch objektiv abweichen. Vermeintliche Konformität (Typ II) liegt – gemittelt über alle 109 Kriterien – bei neun Prozent der Stichprobe vor, vermeintliche Abweichung (oder unwissentliche Konformität; Typ III) bei nur mehr vier Prozent. D.h., daß die nach Fisch (1977) für Interaktionen in Gruppen problematischen sozialen Orientierungslagen der vermeintlichen Konformität (mit den dann auftretenden Schwierigkeiten, wenn für die Gruppe gehandelt werden soll/möchte) und der unwissentlichen Konformität (mit der Gefahr eigentlich unnötiger Kräfteverluste) bei der Bewertung der Evaluationskriterien im Mittel zusammen nur bei 13 Prozent der Stichprobe vorliegen. Insgesamt ergibt sich damit – bei der Analyse über alle 109 Evaluationskriterien, deren soziale Norm inferenzstatistisch abgesichert werden konnte – das Bild einer sozial relativ urteilshomogenen, wohl integrierten Gruppe: Die deutliche Mehrheit der Befragten nimmt die auf die Bewertung von Evaluationskriterien für die eigene berufliche Tätigkeit bezogene Mehrheitsmeinung in der Kollegenschaft richtig wahr und stimmt – im Mittel zu drei Viertel der Stichprobe – mit dieser Mehrheitsmeinung subjektiv und objektiv überein.

Bezogen auf die einzelnen Evaluationskriterien zeigen sich freilich deutliche Unterschiede in den Häufigkeitsverteilungen der Konformitätstypen und -indices (siehe Tabelle 5). So liegen etwa bei den Kriterien der Anzahl der erhaltenen Buchrezensionen, der Beiträge in Kongreß- oder Tagungsbänden, des Renommées von Verlagen, der Editorials, der Diskussionsbeiträge und Repliken, der Edition von Newslettern, der Anzahl betreuter Diplomarbeiten, der nationalen und internationalen Posterpräsentationen, der Berufung in Gremien berufsständischer Organisationen, der Honorarprofessuren, der Lehrstuhlvertretungen, der Funktionen in der akademischen Selbstverwaltung, der Gutachtertätigkeit für Buchmanuskripte, Stipendien und Kongreßbeitragsanmeldungen sowie der Funktionen im BDP die Quoten für den vermeintlich konformen Typus II erheblich höher als im Durchschnitt (nämlich zwischen 13 und 43 Prozent der Stichprobe). Ähnliches gilt für eine erhöhte relative Häufigkeit (auf 13 bis 28 Prozent) des unwissentlich konformen Typus III bei den Evaluationskriterien der Bei-

träge in Kongreß- und Tagungsbänden, der Editorials, der Anzahl betreuter Diplomarbeiten, der internationalen Posterpräsentationen, der auf die Lehre bezogenen Aspekte, der Mitgliedschaft in universitären Gremien sowie der Funktionen im BDP. Auch bewußtes Außenseitertum (im Sinne des Konformitätstyps IV) liegt mit 14 bis 27 Prozent bei ausgewählten Kriterien (wie der Alleinautorenschaft, der Coautorenschaft renommierter Kollegen, dem Verlagsrenommee, der Edition von Buchreihen, dem Qualifikationsalter, den Landes-Drittmittelwerbungen, den nationalen Posterpräsentationen, den Berufungen in politische Expertengremien und in Gremien berufsständischer Organisationen, der Ehrendoktorwürde, der Honorarprofessur, des Fachsprechers/Direktors, des Dekanats sowie der Funktionen in der DGPs) erheblich häufiger vor als in der Mittelung über alle Kriterien.

Exploration von Möglichkeiten zur Aggregation der konsensuellen Evaluationskriterien

Explorativ wurde anhand der persönlichen Bewertungen der 100 Evaluationskriterien, für die in der Stichprobe eine soziale Norm im Sinne einer „sehr positiven Bewertung“ (siehe Tabelle 5) inferenzstatistisch abgesichert werden konnte (für die somit ein hoher positiver Konsens besteht), den Fragen nach den Möglichkeiten ihrer Dimensionalisierung und Aggregation nachgegangen. Die Faktorisierung (Hauptkomponentenanalyse mit Varimax-Rotation) der persönlichen Bewertungen dieser 100 Kriterien führt nach dem Scree-Test zu einer dreifaktoriellen Lösung, durch die 40.1 Prozent der Gesamtvarianz gebunden wird. Da die Mehrzahl der Items Doppel- oder Tripelladungen auf den drei Faktoren aufwies und die Faktorenstruktur somit unrein und (nach dem Fürntratt-Kriterium; vgl. Fürntratt, 1969) kaum zu interpretieren blieb, wurde nach einer sukzessiven Erweiterung des Scree-Tests um einen weiteren Schritt die dann resultierende sechsfaktorielle Lösung (aufgeklärte Varianz: 50.5 Prozent) berechnet. Zur Interpretation der Faktoren wurden ausschließlich reine Faktorladungen ($a_{ij} > .50$) unter Beachtung des Fürntratt-Kriteriums herangezogen (auf die Wiedergabe der Faktormatrix wird hier aus Platzgründen verzichtet; sie wird auf Anfrage selbstverständlich gerne zur Verfügung gestellt):

Faktor I (Eigenwert $e_1 = 28.9$; relativer Varianzanteil: 57.2 %) kann als *Evaluationsdimension der Ämter und Berufungen (anerkanntes Expertentum)* bezeichnet werden. Er verfügt über elf Markieritems, die sich auf erhaltene Ehrendoktorwürden ($a_{ij} = .72$), Berufungen in wissenschaftliche Beiträge (.67), Akademien (.65), Gremien von Wissenschaftsorganisationen (.64), Gremien berufsständischer Organisationen (.61), erhaltene wissenschaftliche Preise (.56), die Funktionen des Fachsprechers/Institutsdirektors (.58) und des Dekans (.54), die Gutachtertätigkeit für Buchmanuskripte (.53), für Stipendiansanträge (.54) und die DFG (.64) beziehen.

Faktor II ($e_2 = 6.5$; 12.8 %) betrifft die *Evaluationsdimension der Anzahl und Wirkung von (Zeitschriften-) Publikationen*. Markiertvariablen sind die Anzahl der Publikationen ($a_{ij} = .63$), die Anzahl der Fremdzitationen (.69), die Anzahl der „cited core documents“ (.57), der Rang der zitierenden Publikationsquellen (.70), das Renommee der Zeitschrift (.71) und die Veröffentlichung experimenteller Untersuchungen (.58).

Faktor III ($e_3 = 4.7$; 9.3 %) betrifft die *Evaluationsdimension der Kongreß- und Tagungsaktivitäten* (national und international). Seine Markiertvariablen sind die Anzahl der angemeldeten Hauptvorträge ($a_{ij} = .78$), eingeladenen Hauptvorträge (.69), Überblicks-/Positionsreferate (.73), Forschungsreferate (.74) und Organisationen von Arbeitsgruppen (.66) auf nationalen Kongressen und Tagungen sowie die Anzahl der angemeldeten Hauptvorträge ($a_{ij} = .78$), eingeladenen Hauptvorträge (.64), Überblicks-/Positionsreferate (.76), Forschungsreferate (.78), Organisationen von Arbeitsgruppen (.66) und Posterpräsentationen (.59) auf internationalen Kongressen und Tagungen.

Faktor IV ($e_4 = 3.8$; 7.6 %) betrifft die *Evaluationsdimension der Buch- und Testpublikationen sowie Buchbeiträge*. Seine Markiertvariablen sind die Anzahl der publizierten Monographien ($a_{ij} = .55$), Buchbeiträge (.61), Lehrbücher (.57), Enzyklopädiebeiträge (.55), Testmanuale (.50) und theoretischen Analysen (.53) sowie die Anzahl der Herausgeberschaften von Sammelwerken (.52).

Faktor V ($e_5 = 3.6$; 7.2 %) betrifft die *Evaluationsdimension der Höhe der eingeworbenen Drittmittel*. Seine Markiertvariablen beziehen sich auf Drittmittelinwerbungen vom Bund ($a_{ij} = .80$), von den Ländern (.78), von der Europäischen Union (.77), von Stiftungen (.77), von der Wirtschaft (.71) und von sonstigen Quellen (.72). Anzumerken ist, daß die von der DFG eingeworbenen Drittmittel auf diesem Faktor zwar zu $a_{ij} = .49$ laden, wegen hoher Simultanladungen auf Faktor I (.33) und Faktor II (.54) aber nicht dem Fürntratt-Kriterium genügen.

Faktor VI ($e_6 = 2.9$; 5.8 %) betrifft die *Evaluationsdimension des Engagements in der Lehre und ihrer Qualität*. Seine Markiertvariablen beziehen sich auf die Anzahl der betreuten Diplomarbeiten ($a_{ij} = .63$) und Dissertationen (.59), das Engagement in der Lehre (.73), die Breite der Lehrerfahrung (.68), die didaktische Aufbereitung und Qualität der Lehrveranstaltungen (.75), die individuelle Betreuung und Beratung von Studierenden (.76), die Praxis- und Anwendungsorientierung der Lehre (.57) sowie die Aktualität der Lehrveranstaltungen (.64).

Die Faktorisierung der persönlichen Bewertungen der 100 in der Stichprobe konsensuell als sehr positiv bewerteten Evaluationskriterien hat somit zu einer relativ gut interpretierbaren Struktur von Evaluationsdimensionen geführt. Auffällig ist dabei, daß sich für ganze Gruppen der vorgegebenen Evaluationskriterien aufgrund von Doppelladungen oder aufgrund von Faktorladungen, die nach dem Fürntratt-Kriterium nicht bedeutsam sind, keine klaren Zuordnungen ergeben haben. In der Faktor-

lösung nicht bzw. schlecht (unrein) repräsentiert sind dabei vor allem Kriterien aus den Themenbereichen (1.) der Qualität wissenschaftlicher Arbeiten (mit keinen essentiellen Faktorladungen!), (2.) des Qualifikationsalters (mit Doppelladungen auf den Faktoren I und II), (3.) der Anzahl der Forschungsprojekte, -mitarbeiter und -kooperationen (mit Mehrfachladungen auf den Faktoren I, II, III und V), (4.) der Auslandsaufenthalte und Reputationseinstufungen durch Kollegen (mit Mehrfachladungen auf den Faktoren I, II und V) sowie (5.) der Funktionen in Fachgesellschaften (mit Doppelladungen auf den Faktoren I und III).

Auffällig bei dieser Faktorenlösung ist zum zweiten, daß sich die meisten Markiertvariablen der Faktoren auf direkt und einfach quantifizierbare Evaluationskriterien oder zumindest auf solche beziehen, die prinzipiell einer Quantifizierung zugänglich sind (wie z.B. der Rang zitierender Quellen über den „impact“-Faktor von Zeitschriften oder das Engagement in der Lehre über Deputatsrealisierungen und/oder Hörerbefragungen etc.). Quantitativ nicht oder erheblich schwerer zugängliche, mithin qualitative Evaluationskriterien (wie etwa alle Bewertungsaspekte für die Qualität wissenschaftlicher Arbeiten, das Renommee zitierender Autoren, der Rang berufender Institutionen, Reputationseinstufungen durch Kollegen etc.) weisen dagegen keine essentiellen Ladungen oder Mehrfachladungen in der Faktorenlösung auf. Damit ist freilich keinerlei Aussage über eine geringere Relevanz qualitativer Evaluationskriterien verbunden, da ja auch diese in der Faktorenlösung nicht oder schlecht repräsentierten qualitativen Evaluationskriterien von der inferenzstatistisch abgesicherten Mehrheit der Befragten als „sehr positiv“ bewertet wurden. Die vorgestellte Faktorenlösung und die darauf basierenden Explorationen von Möglichkeiten der Bildung psychometrischer Subskalen zu Bewertungsdimensionen von Evaluationskriterien, die im folgenden dargestellt werden, beziehen sich also nur auf einen (den quantitativen) Aspekt der Evaluation der Tätigkeiten von Hochschullehrern(innen), der auf jeden Fall der Ergänzung um den qualitativen Aspekt bedarf (siehe hierzu auch die Ausführungen in der Diskussion).

Die oben aufgeführten Markiertvariablen der sechs Faktoren zu Bewertungsdimensionen wurden als Grundlage für die Bildung von *sechs Skalen zu Bewertungsdimensionen für (quantitative/quantifizierbare) Evaluationskriterien* (SBEV) verwendet. Lediglich Skala V (Drittmittelinwerbungen) wurde aus inhaltlichen Gründen um das Item zur Bewertung von Drittmittelinwerbungen bei der DFG ergänzt, obwohl es keine reine Ladung auf Faktor V aufweist (siehe oben). In Tabelle 6 sind die deskriptiven Parameter dieser sechs Skalen, ihre Interkorrelationen und ihre Korrelationen mit ausgewählten objektiven Variablen aufgeführt. Vorläufig können die Skalen wie folgt benannt werden:

- SBEV-I: Evaluationsdimension der Ämter und Berufungen (anerkanntes Expertentum)
- SBEV-II: Evaluationsdimension der Anzahl und Wirkung von (Zeitschriften-)Publikationen

- SBEv-III: Evaluationsdimension der Kongreß- und Tagungsaktivitäten
- SBEv-IV: Evaluationsdimension der Buch- und Testpublikationen sowie Buchbeiträge
- SBEv-V: Evaluationsdimension der Höhe der eingeworbenen Drittmittel
- SBEv-VI: Evaluationsdimension des Engagements in der Lehre und ihrer Qualität.

Bei vergleichbaren Standardabweichungen (lediglich Skala V weist hier einen etwas höheren Wert, mithin eine erhöhte Urteilsheterogenität in der Stichprobe, auf) liegen die anhand der jeweiligen Itemzahl relativierten Skalenmittelwerte alle im deutlich positiven Bereich, was schlicht und einfach aus dem Vorgehen bei der Itemselektion (Auswahl der 100 konsensuell als sehr positiv bewerteten Evaluationskriterien) resultiert. Ihre Rangreihe weist darauf hin, daß Evaluationskriterien aus den Bereichen der Zeitschriftenpublikationen und der Lehr-tätigkeit konsensuell am positivsten bewertet werden. Ihnen folgen Indikatoren aus den Bereichen der Kongreß/ Tagungsaktivitäten, der Buch- und Beitragspublikationen, der Drittmittelinwerbungen und des anerkannten Expertentums (siehe Tabelle 6). Die interne Konsistenz (nach Cronbachs Alpha) aller sechs Skalen zur Bewer-

tung von Evaluationsdimensionen liegt im oberen Bereich, was durch hohe, statistisch signifikante Trennschärfekoeffizienten aller Items fundiert wird. Mit einer Ausnahme werden alle Skaleninterkorrelationen signifikant, was – trotz Skalenableitung nach den Befunden einer Faktorenanalyse nach der Hauptkomponentenmethode mit Varimax-Rotation – kaum verwundert, da alle Items in der Stichprobe als konsensuell sehr positiv bewertet wurden. Wichtiger ist dagegen die Frage nach der Reliabilität der Skalendifferenzen bzw. der Profilreliabilität aller sechs Skalen, da sie eine Aussage darüber gestattet, ob es empirisch gelungen ist, unterschiedliche Aspekte in der Bewertung von Evaluationsdimensionen mit diesen sechs Skalen hinreichend differenziert und dabei reliabel abzubilden. Die anhand der internen Konsistenzen und Skaleninterkorrelationen (siehe Tabelle 6) berechnete Profilreliabilität (vgl. etwa Lienert, 1969) der SBEv beträgt $\text{prof}r_{tt} = .81$ und bestätigt, daß eine differenzierte Erfassung der Bewertung von sechs unterschiedlichen Evaluationsdimensionen hinreichend reliabel erreicht wurde.

In Tabelle 6 finden sich ebenfalls die Korrelationen der sechs Skalen mit ausgewählten Referenzvariablen (siehe auch Tabelle 4): Es zeigen sich keine bedeutsamen Zusammenhänge zwischen den Skalenwerten und der

Tabelle 6. Deskriptive Parameter, Interkorrelationen und Außenkorrelationen der faktorenanalytisch fundierten Skalen zu Bewertungsdimensionen für Evaluationskriterien (SBEv)

Skalenparameter	SBEv-I	SBEv-II	SBEv-III	SBEv-IV	SBEv-V	SBEv-VI
Itemanzahl	11	7	11	7	7	8
relativierter Mittelwert (<i>M</i>)	5.0	6.0	5.5	5.5	5.3	5.9
Standardabweichung (<i>SD</i>)	1.3	1.2	1.3	1.0	1.6	1.1
interne Konsistenz (α)	.90	.84	.94	.78	.95	.87
Trennschärfekoeff. (Range r_{it})	.57–.74	.50–.75	.59–.83	.30–.68	.66–.90	.50–.77
Skaleninterkorrelationen (<i>r</i>)						
SBEv-I: Ämter und Berufungen	1.00	.30**	.56**	.47**	.51**	.49**
SBEv-II: (Zeitschriften-)Publik.		1.00	.28**	.12	.44**	.17**
SBEv-III: Kongreßaktivitäten			1.00	.44**	.42**	.37**
SBEv-IV: Bücher und Buchbeiträge				1.00	.33**	.45**
SBEv-V: Drittmittelinwerbung					1.00	.41**
SBEv-VI: Lehre						1.00
Referenzvariable (<i>r</i>)						
PSYNDEX-Gesamtnachweise	.07	.05	-.09	-.02	.03	-.09
PsycLit-Gesamtnachweise	-.02	.19**	-.12*	-.18**	-.01	-.21**
SSCI-Fremdzitationen	-.04	.18**	-.10	-.11	.04	-.10
Anzahl betreuter Dissertationen	.05	.08	.05	.01	.02	-.01
Alter	.05	-.10	.05	.17**	-.11	-.07
Größe des Heimatinstituts:						
– Anzahl der Studierenden	-.09	.09	-.15*	.04	-.07	.02
– Anzahl der Professuren	-.03	.05	-.13*	.03	-.17**	-.06

Anmerkungen: ** $p < .01$; * $p < .05$

Gesamtzahl der Publikationsnachweise im PSYINDEX sowie der Zahl der betreuten Dissertationen. Für die Gesamtzahl der Publikationsnachweise in PsycLit, die Anzahl der Fremdzitationen nach dem SSCI, das Alter und die Größe des Heimatinstituts liegen dagegen einige statistisch bedeutsame, numerisch jedoch niedrige Korrelationskoeffizienten vor (maximale Varianzbindung von 2 %; siehe Tabelle 6). Zu ergänzen ist, daß weder für das Dienstalster (Zeit seit der Erstberufung bzw. Habilitation: $r < |.12|$) noch für die Anzahl der nach der Erstberufung erhaltenen Rufe ($r < |.13|$) signifikante Korrelationen zu den Skalenwerten vorliegen. Insgesamt kann jedoch davon ausgegangen werden, daß die persönliche Bewertung der sechs Evaluationsdimensionen in der Stichprobe nicht in hohem Maße mit diesen Referenzvariablen kovariiert (und darauf bezogene Urteilsverzerrungen vorliegen). In Vergleichen der Geschlechtsgruppen ergaben sich dagegen Hinweise darauf, daß Kolleginnen auf Skala III ($M_w = 5.9, M_m = 5.4; t(256) = 2.36, p < .05$), Skala IV ($M_w = 5.9, M_m = 5.4; t(255) = 2.80, p < .01$) und Skala VI ($M_w = 6.3, M_m = 5.8; t(256) = 2.36, p < .05$) im Durchschnitt signifikant höhere Werte aufweisen als die Kollegen (alle anderen Skalen: $t(253) < 1.61$). Für den Status (Prof. versus Priv.-Doz.) zeigt sich dagegen lediglich auf Skala VI ein signifikanter Gruppenunterschied ($M_{\text{prof}} = 5.8, M_{\text{pd}} = 6.3; t(215) = 2.61, p < .01$; alle anderen Skalen: $t(215 < 1.45)$).

Diskussion

Die Befunde der Expertenbefragung führen zunächst zur Bejahung der empirischen Ausgangsfrage: Für die Gruppe der an deutschen Psychologischen Instituten mit Hauptfachausbildung tätigen Hochschullehrer(innen) ist es gelungen, eine relativ konsensuelle Liste von Evaluationskriterien für eigene beruflichen Leistungen im universitären Kontext zu identifizieren. Für immerhin 109 von 117 zur Wichtigkeitsbewertung vorgegebenen, prinzipiell möglichen Evaluationskriterien konnte inferenzstatistisch eine klare soziale Norm abgesichert werden. Lediglich für acht der vorgegeben Kriterien ließen sich somit keine klaren statistischen Gruppennormen identifizieren, d. h., daß bei diesen Kriterien (Erwähnungen in den Medien, Publikation „grauer“ Literatur und in Newslettern sowie Veröffentlichung von Multimedia Produkten, Softwareprodukten, Buchbesprechungen, Schulungs-/Trainingsmaterialien und Kongreß-/Tagungsberichten) die Meinungen geteilt sind. Von der deutlichen Mehrheit als „weniger positiv“ bewertet wurden nur neun der vorgegebenen Kriterien (nämlich die Anzahl der Rezensionen eigener Bücher, der Publikation von Beiträgen in Kongreß- oder Tagungsbänden, der Software-Besprechungen, der Editorials, der Herausgabe von Newslettern, der Präsentation von Postern auf nationalen wissenschaftlichen Kongressen oder Tagungen, der Mitgliedschaft im Fachbereichs- oder Fakultätsrat, der Mitarbeit in anderen universitären Gremien der Selbstverwaltung sowie der Funktionen im BDP). Zusammengekommen weisen diese Evaluationskriterien mit unklarer

oder weniger positiver Gruppennorm dadurch auf eine berufsgruppenspezifische Problematik, daß sie sich vor allem auf Aspekte der Öffentlichkeitsarbeit, der fachinternen und -externen Dienstleistungen sowie der (Vorab-)Veröffentlichung aktueller Forschungsergebnisse beziehen, die eigentlich für ein Fach bedeutsam sind. Gleichwohl werden entsprechende Arbeitsleistungen in der Bezugsgruppe der Kollegen und Kolleginnen weniger anerkannt bzw. höchst unterschiedlich (d. h., sozial unklar) bewertet, was etwa die von Herrmann (1996) geäußerten Bedenken an einer zunehmenden persönlichen Profitorientierung im Wissenschafts-Selbstverständnis der modernen akademischen Psychologie (mit primär an individuellen Vorteilen orientierten Verkaufsmotivationen „nach dem Muster des Gemüsehandels“; Herrmann, 1996, S. 216) bestärkt. Abgeschwächt werden diese Bedenken jedoch dadurch, daß unter den verbleibenden 100 Evaluationskriterien für die Tätigkeit psychologischer Hochschullehrer(innen) mit einer eindeutigen, sehr positiven sozialen Norm sich auch solche finden, die fachinterne und -externe Dienstleistungen betreffen.

Unter Bezug auf die Ausgangsfrage ist somit festzuhalten, daß für 93 Prozent der vorgegebenen Kriterien in der Berufsgruppe der Hochschullehrer(innen) an Psychologischen Instituten mit Hauptfachausbildung ein Konsens bei der Bewertung besteht. Dabei ergibt sich unter Bezug auf die KTA-Kriterien der Kenntnis der sozialen Normen und der Konformitätstypen (wissentliche Konformität und wissentliche Nonkonformität) ferner insgesamt das Gesamtbild einer gut integrierten Berufsgruppe mit hohem Konsensus bei der Wichtigkeitseinschätzung von Kriterien zur Evaluation eigener Arbeitsleistungen. Während Fisch (1977) – ebenfalls mit der Methodik der KTA – in einer ähnlichen großen Stichprobe psychologischer Hochschullehrer(innen) für Beurteilungen der persönlichen Situation, der täglichen Arbeit in Forschung und Lehre sowie der Hauptziele der beruflichen Tätigkeit „ein Spektrum von Orientierungsschwierigkeiten“ zu konstatieren hatte, da es „zum einen (...) an Übereinstimmung mit den Fachkollegen (mangelt), zum anderen an einer zutreffenden Lageeinschätzung der eigenen Fachkollegen“ (Fisch, 1977, S. 154–155), liegt nach den hier präsentierten Daten gemittelt über die konsensuell bewerteten Kriterien eine gute kognitive Orientierung hinsichtlich berufsspezifischer Bewertungsmaßstäbe für Arbeitsleistungen in psychologischer Forschung und Lehre vor. Dokumentiert wurden dabei freilich auch Unterschiede in den auf verschiedene Evaluationskriterien bezogenen sozialen Orientierungslagen, die ebenso wie die von den Befragten frei ergänzten Evaluationskriterien Anlaß zu Folgeuntersuchungen sowie zu ihrer Beachtung in Evaluationsmaßnahmen und -studien geben. Dies gilt vor allem für die Aspekte des Theorienbezugs empirischer Arbeiten, ihrer Interdisziplinarität bzw. Bedeutung für andere Fächer sowie ihrer praktischen und gesellschaftlichen Relevanz.

Mithin ist berufsgruppen-intern eine konsensuelle Validierung einer umfangreichen Liste tätigkeitsrelevanter Evaluationskriterien gelungen, die etwa auch extern als Argumentationshilfe (etwa bei externen Hemmnissen

und Schwierigkeiten bei der Durchsetzung adäquater Arbeitsbedingungen; vgl. hierzu etwa Fisch & Daniel, 1981) genutzt werden kann. Insbesondere kann mit Hilfe der vorgelegten Befunde solchen Evaluationsbestrebungen im Rahmen interner oder externer Maßnahmen der Qualitätssicherung oder des Qualitätsmanagements entgegen werden, die auf singuläre, auf wenige oder auf für die Psychologie nach der Mehrheitsmeinung der von uns befragten Experten weniger geeignete Evaluationskriterien setzen. Die vorgelegten Ergebnisse weisen auf die Notwendigkeit der Verwendung vieler Kriterienvariablen bei der Evaluation der Arbeitsleistungen von Hochschullehrern(innen) der Psychologie und machen die Anwendung singulärer oder weniger Kriterienvariablen, die der Komplexität und Breite des Tätigkeitsfeldes nicht gerecht werden, empirisch obsolet. Unter Bezug auf die Diskussion um die internationale Dissemination psychologischer Befunde und Theorien aus dem deutschsprachigen Bereich (vgl. etwa Keul et al., 1993, 1994; Montada et al., 1995) sei hier noch einmal darauf verwiesen, daß bei allen Evaluationskriterien, die separat nach nationalen versus internationalen Aktivitäten (wie Kongreß- und Tagungsbesuche, wissenschaftliche Kooperationen, Gastprofessuren sowie Funktion in Fachgesellschaften) zu bewerten waren, die internationalen Aktivitäten durchgängig höhere Mittelwerte erzielten, diesen mithin ein besonderer, konsensuell hoher Stellenwert zukommt.

Auf dem Hintergrund der hohen Anzahl der konsensuell von den Befragten als positiv bewerteten Evaluationskriterien stellen sich die Fragen nach ihren Inhalten und ihrer Datenmodalitäten. Faktorenanalytisch wurde dazu explorativ den Möglichkeiten der Dimensionierung der Kriterien auf der Ebene ihrer Bewertung durch Experten (in eigener Sache) nachgegangen. Für die konsensuell als sehr positiv bewerteten 100 Kriterien ergab sich eine Faktorenlösung, aus der sechs Subskalen zu Bewertungsdimensionen für quantitative oder quantifizierbare Evaluationskriterien entwickelt werden konnten. Diese Subskalen erwiesen sich als hoch reliabel, sie verfügen zudem über eine gute Profilreliabilität, und genügen nach dem Aspekt der konsensuellen Validierung durch Experten auch dem Kriterium der Konstruktvalidität. Auffällig ist jedoch, daß nahezu ausschließlich solche Evaluationskriterien Markiertvariablen der Faktoren sind (und mithin in die Skalenbildung Eingang gefunden haben), die direkt quantifizierbar oder zumindest prinzipiell einer Quantifizierung zugänglich sind. Es scheint, als ob die quantitativ-statistische Analyse (der Faktorenanalyse) für quantitative Daten (persönliche Wichtigkeitsbeurteilungen) hier nicht dazu taugt, qualitative Kriterienvariablen aufzunehmen bzw. „rein“ zu repräsentieren (für die die Faktorenanalyse ja auch nicht entwickelt wurde und nicht/weniger geeignet ist). Damit ist die der Psychologie historisch und auch aktuell sehr bekannte Frage nach quantitativem und qualitativem Vorgehen (bzw. entsprechenden Variablen) sowie das darauf bezogene methodologische Spannungsverhältnis angesprochen, das auch heute noch allzu häufig mit einem „Entweder-Oder“ beantwortet wird und aufgelöst werden will.

In der internationalen Fachliteratur zur Wissenschaftsforschung und -evaluation ist es seit Jahren konzeptuell und in der praktischen Umsetzung eine Binneweisheit, daß sich qualitatives und quantitatives Vorgehen zu ergänzen haben (siehe etwa im Überblick Chubin & Hackett, 1990; Daniel, 1996; Feld et al., 1995; van Raan, 1996 a, 1996 b). Einseitige Vorgehensweisen sind obsolet, da quantitativen Evaluationskriterien häufig zumindest fragliche Validitäten für das Konzept der Produktivität und Kreativität in Lehre und Forschung zugeschrieben werden (vgl. hierzu etwa Frühwald, 1997; Herrmann, 1996; Süllwold, 1997 b) und da qualitative Evaluationskriterien in der Regel erhebliche Defizite in den Bereichen der Objektivität und Reliabilität aufweisen. Quantifizierungsversuche weisen exakt hier Vorteile auf, die ebenso wie in anderen Forschungsbereichen der Psychologie zunächst genutzt und dann empirisch im Hinblick auf ihre Validität überprüft werden sollten. Nicht nur in der psychologischen Kreativitätsforschung (vgl. hierzu etwa Krampen, 1993), sondern aller Orten finden sich in der Psychologie Quantifizierungs- und Psychometrisierungsversuche für operational schwer zugängliche Konzepte. Das, was dort konzeptgeleitet empirisch versucht wird, sollte auch dann getan werden, wenn es um die eigenen Arbeitsleistungen geht. Freilich sind vergleichende Prüfungen der Güte quantitativ und qualitativ gewonnener Variablen sowie ihre wechselseitige konkurrente Validierung notwendig. Die (vorläufigen) Lösungsversuche für das Dilemma zwischen quantitativem und qualitativem Zugang betonen daher auch die Notwendigkeiten des kombinierten Vorgehens, wobei die qualitativen Aspekte zumeist über „peer-review“-Verfahren und die quantitativen Aspekte zumeist über scientometrische Methoden und Hörerbefragungen umgesetzt werden. Dafür und für einiges mehr liefert die hier vorgelegte konsensuelle Liste quantitativer, prinzipiell quantifizierbarer und qualitativer Evaluationskriterien für die Tätigkeit von Hochschullehrern(innen) im Fach Psychologie ein breites Reservoir, das in Folgearbeiten sowie bei Evaluationsmaßnahmen auf Personen- und auf Instituts-ebene in der Psychologie genutzt werden kann.

In der Einleitung wurde auf die große Bedeutung der Akzeptanz von Evaluationsmaßnahmen und -kriterien bei den davon Betroffenen verwiesen. Hier wurde versucht, eine möglichst breite Akzeptanz dadurch zu erreichen, daß die Entwicklung der Kriterien intern (von der eigenen Berufsgruppe ausgehend) vorgenommen wurde und nur solche Kriterien als gültig akzeptiert wurden, die von der großen (inferenzstatistisch abgesicherten) Mehrheit der Kollegenschaft als „sehr positiv“ bewertet wurden. Damit ist die Repräsentativität der Stichprobe von zentraler Bedeutung. Zunächst ist zu betonen, daß sich die Untersuchungspopulation ausschließlich auf Professoren(innen) und Privat-Dozenten(innen) an deutschen Psychologischen Universitätsinstituten mit Hauptfachausbildung bezieht. Die Befunde dürfen daher nicht ohne weiteres auf andere Personenkreise (etwa wissenschaftliche Mitarbeiter, Österreicher, Schweizer, Hochschullehrer an Instituten mit Nebenfachausbildung etc.) generalisiert werden. Ferner war die Repräsentativität der

Stichprobe (der Fragebogen-Beantworter) zu prüfen. Dies wurde für eine Vielzahl objektiv gewonnener Variablen zu Person- und Arbeitsmerkmalen realisiert. Dabei konnte die Repräsentativität der Stichprobe nach den Kriterien der Geschlechtsverteilung, des beruflichen Status (Prof. vs Priv.-Doz.), der Größe des Heimatinstituts, der Anzahl der Fremd- und Selbstzitationen nach dem SSCI sowie den PSYNDEX-Nachweiszahlen für Bücher, Buchbeiträge und „graue“ Veröffentlichungen (Instituts- und Arbeitsberichte) belegt werden. Repräsentativitätseinschränkungen liegen bei den Nachweisanzahlen der Zeitschriftbeiträge (nach PSYNDEX, PsycLit und SSCI) sowie der Bücher und Buchbeiträge (nach PsycLit) vor, da die Fragebogen-Beantworter hier im Durchschnitt über signifikant höhere Werte verfügen als die Nichtantworter. Analoges gilt für die Anzahl der betreuten Dissertationen, für die die Nichtantworter im Mittel einen signifikant höheren Wert aufweisen als die Fragebogen-Beantworter. Diese Repräsentativitätseinschränkungen konnten jedoch durch Analysen der Zusammenhänge dieser kritischen Variablen mit den persönlichen Bewertungen der darauf bezogenen Evaluationskriterien innerhalb der Stichprobe der Fragebogen-Beantworter relativiert werden. Es zeigten sich entweder keine oder nur schwach ausgeprägte Zusammenhänge zwischen den objektiven Indikatoren und den darauf bezogenen persönlichen Bewertungen, so daß systematische Urteilsverzerrungen unwahrscheinlich sind bzw. gering sein werden. Anhand des vorliegenden Datenmaterials können natürlich systematische Interaktionen zwischen den Gründen für die Untersuchungsteilnahme versus die Nichtteilnahme und dem Untersuchungsgegenstand nicht ausgeschlossen werden.

Literatur

- Alewell, K. (1988). Beurteilung der Leistungen von Hochschulen. In H.-D. Daniel & R. Fisch (Hrsg.), *Evaluation von Forschung* (S. 41–58). Konstanz: Universitätsverlag Konstanz.
- Bolsenkötter, H. (1986). Ansätze zur Erfassung und Beurteilung von Forschungsleistungen. In R. Fisch & H.-D. Daniel (Hrsg.), *Messung und Förderung von Forschungsleistung* (S. 41–50). Konstanz: Universitätsverlag Konstanz.
- Brandstädter, J. & Krampen, G. (1979). Objektive und subjektive Konformität von Lehrern in der Bewertung pädagogischer Zielorientierungen. *Psychologie in Erziehung und Unterricht*, 26, 140–148.
- Breznitz, S. (1967). Confidence estimation of group norm as a function of subjective conformity. *Psychonomic Science*, 7, 399–400.
- Chubin, D. E. & Hackett, E. J. (1990). *Peerless science, peer review and US science policy*. New York, NY: State University of New York Press.
- Daniel, H.-D. (1986). Die Vermessung der Forschung. In H. Methner (Hrsg.), *Psychologie in Betrieb und Verwaltung* (S. 208–218). Bonn: Deutscher Psychologen Verlag.
- Daniel, H.-D. (1988). Forschungsleistungen wissenschaftlicher Hochschulen im Vergleich. In H.-D. Daniel & R. Fisch (Hrsg.), *Evaluation von Forschung* (S. 93–106). Konstanz: Universitätsverlag Konstanz.
- Daniel, H.-D. (1994). Hörerbefragung an der Universität Mannheim. *Empirische Pädagogik*, 8, 109–129.
- Daniel, H.-D. (1996, September). *Kriterien für die Evaluation von Forschungsleistungen: Konzept für eine Delphi-Befragung der Mitglieder der DGPs*. Vortrag auf dem 40. Kongreß der Deutschen Gesellschaft für Psychologie in München.
- Daniel, H.-D. & Fisch, R. (Hrsg.). (1988). *Evaluation von Forschung*. Konstanz: Universitätsverlag Konstanz.
- Daniel, H.-D., Frühwald, W., Schiedermaier, H. & Erichsen, H.-U. (1995). Hochschul-Ranking: Gibt es die „beste Uni“? *Universitas*, 50 (3), 205–220.
- Deutsche Gesellschaft für Psychologie (1997). Differenzierte Zuweisung von Haushaltsmitteln an Hochschulen: Ergebnisse und Anmerkungen zu einer Umfrage des DGPs-Vorstandes an Psychologischen Instituten. *Psychologische Rundschau*, 48, 163–166.
- Endler, N. S., Rushton, J. P. & Roediger, H. L. (1978). Productivity and scholarly impact (citations) of British, Canadian, and U.S. Departments of Psychology (1975). *American Psychologist*, 33, 1064–1082.
- Felt, U., Nowotny, H. & Taschwer, K. (1995). *Wissenschaftsforschung*. Frankfurt/Main: Campus.
- Fisch, R. (1977). Aspekte sozialer Orientierung bei Wissenschaftlern: Untersucht am Beispiel des Fachs Psychologie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 29, 137–156.
- Fisch, R. & Daniel, H.-D. (1981). Die Lage der Forschung im Fach Psychologie im Urteil der Kollegen. *Psychologische Rundschau*, 32, 103–117.
- Fisch, R. & Daniel, H.-D. (Hrsg.). (1986). *Messung und Förderung von Forschungsleistung*. Konstanz: Universitätsverlag Konstanz.
- Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, 44, 155–194.
- Frühwald, W. (1997). Sinnloser Datensalat. Interview mit DFG-Präsident W. Frühwald. *DIE ZEIT*, 29, S. 63.
- Fürntratt, E. (1969). Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in Faktorenanalysen psychologischer Daten. *Diagnostica*, 15, 62–75.
- Heckhausen, H. (1988). Wozu ortsvergleichende Produktivitäts-Ranglisten von Fachbereichen und wie deshalb vorgegangen werden sollte. In H.-D. Daniel & R. Fisch (Hrsg.), *Evaluation von Forschung* (S. 33–40). Konstanz: Universitätsverlag Konstanz.
- Hellstern, G.-H. (1997, Oktober). *Die Notwendigkeit für eine Evaluierung als organisatorischer Lernprozeß*. Vortrag auf dem 20. Kongreß der Deutschen Vereinigung für Politische Wissenschaften in Bamberg, Arbeitsgruppe „Hochschulevaluierung“.
- Herrmann, T. (1982). Über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. *Sprache und Kognition*, 1, 3–14.

- Herrmann, T. (1996). Psychologie und Gemüsehandel. *Psychologische Rundschau*, 47, 216–217.
- Hornborstel, S. & Daniel, H.-D. (1994). Das SPIEGEL-Ranking. In P. P. Mohler (Hrsg.), *Universität und Lehre* (S. 29–44). Münster: Waxmann.
- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. London: Sage.
- Keul, A., Gigerenzer, G. & Stroebe, W. (1993). Wie international ist die Psychologie in Deutschland, Österreich und der Schweiz? *Psychologische Rundschau*, 44, 259–269.
- Keul, A., Gigerenzer, G. & Stroebe, W. (1994). Publikationen in internationalen Zeitschriften. *Psychologische Rundschau*, 45, 111–113.
- Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. H. Ingenkamp & R. Jäger (Hrsg.), *Tests und Trends* (= Jahrbuch der Pädagogischen Diagnostik, Bd. 10, S. 11–39). Weinheim: Beltz.
- Krampen, G. (1994). Transkulturelle und ethnopsychologische Analysen sozialer Kognitionen. *Zeitschrift für Politische Psychologie*, 2, 317–336.
- Krampen, G. & Krämer, A. (1995). Objektive und subjektive Konformität in Erklärungen der Ausländerfeindlichkeit in Deutschland. *Zeitschrift für Sozialpsychologie*, 26, 272–284.
- Krampen, G., Freilinger, J. & Wilmes, L. (1994). Entwicklungsbezogene Orientierungen im Umgangswissen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 26, 185–196.
- Krampen, G., Burkard, P., Wiesenhütter, J. & Montada, L. (1998). Psychologische Diplomarbeiten und Dissertationen im deutschsprachigen Bereich. In G. Krampen & H. Zayer (Hrsg.), *Psychologiedidaktik und Evaluation I* (S. 115–136). Bonn: Deutscher Psychologen Verlag.
- Krauss, R. M. & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, 9, 2–24.
- Lienert, G. A. (1969). *Testaufbau und Testanalyse* (3. Aufl.). Weinheim: Beltz.
- Lienert, G. A. (1973). *Verteilungsfreie Methoden in der Biostatistik* (Bd. 1). Meisenheim/Glan: Hain.
- Lonner, W. J. (1968). The SVIB visits German, Austrian, and Swiss psychologists. *American Psychologist*, 23, 164–179.
- Montada, L., Becker, J., Schöpflin, U. & Baltes, P. B. (1995). Die internationale Rezeption der deutschsprachigen Psychologie. *Psychologische Rundschau*, 46, 186–199.
- Mummendey, A. (1996, September). *Einleitung zu: Evaluation psychologischer Institute*. Vortrag auf dem 40. Kongreß der Deutschen Gesellschaft für Psychologie in München.
- Mummendey, A. & Wottawa, H. (1996, September). *Evaluation psychologischer Institute*. Podiumsdiskussion auf dem 40. Kongreß der Deutschen Gesellschaft für Psychologie in München.
- Orlik, P., Fisch, R. & Saterdag, H. (1971). Fragen der sozialen Orientierung von Studienanfängern des Faches Psychologie. *Psychologische Rundschau*, 22, 17–37.
- Plaum, E., Schweitzer, F. & Scharlach, B. (1989/90). Äußerungen von Hochschullehrern zu Vergangenheit, Gegenwart und Zukunft der Psychologie. *Psychologie und Geschichte*, 1 (2), 49–57.
- Rindermann, R. (1998). Skalen der Lehrevaluation: Welche Aspekte sollen in universitären Lehrveranstaltungen beurteilt werden? In G. Krampen & H. Zayer (Hrsg.), *Psychologiedidaktik und Evaluation I* (S. 295–316). Bonn: Deutscher Psychologen Verlag.
- Rindermann, R. & Amelang, M. (1994). *Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE)*. Heidelberg: Asanger.
- Schneider, D. J. (1991). Social cognition. *Annual Review of Psychology*, 42, 527–561.
- Spada, H. (1997). Lage und Entwicklung der Psychologie in Deutschland, Österreich und der Schweiz. *Psychologische Rundschau*, 48, 1–15.
- Süllwold, F. (1997 a). Ranking ist oft ein Synonym für Unsinn. *Forschung & Lehre*, 11/1997, 578–579.
- Süllwold, F. (1997 b). Fehler und Fallen: Anmerkungen zur Evaluation von Forschung und Lehre. *Forschung & Lehre*, 4/1997, 185–188.
- Tack, W. H. (1994). Bericht über Reaktionen auf einen Bericht. *Psychologische Rundschau*, 45, 108–111.
- van Raan, A. F. J. (1993). Advanced bibliometric methods to assess research performance and scientific development. *Research Evaluation*, 3 (3), 151–166.
- van Raan, A. F. J. (1996 a). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, 36, 397–420.
- van Raan, A. F. J. (1996 b, September). *Sinn und Unsinn bei der Anwendung bibliometrischer Methoden für die Messung von Forschungsleistungen im allgemeinen und in der Psychologie im besonderen*. Vortrag auf dem 40. Kongreß der Deutschen Gesellschaft für Psychologie in München.
- Weingart, P. & Winterhager, M. (1984). *Die Vermessung der Forschung*. Frankfurt/Main: Campus.
- Weingart, P., Sehringer, R. & Winterhager, M. (Hrsg.). (1992). *Representations in science and technology*. Leiden: DSWO Press.

Prof. Dr. Leo Montada

Universität Trier
 Fachbereich I – Psychologie
 D-54286 Trier