

11. Fazit

Mit diesem Beitrag der rechtlichen Aspekte der Diplomarbeit drängt sich für den Leser der Eindruck auf, Juristen suhlten sich wie Schwarzkittel nur in pathologischen Fällen. Soviel Problempotential, wie hier auf einmal vereinigt, wird sich bei einer Diplomarbeit, die man in der Regel nur einmal im Leben schreibt, kaum verwirklichen. Die Lektüre mag hoffentlich zusätzlich dazu beitragen, mögliche Problemsituationen vorausschauend zu identifizieren und mit prophylaktischen Strategien zu umgehen. Gefahr erkannt, Gefahr gebannt, gilt auch hierbei.

Quelle:
 * Engel, Stefan & Slapnicar, K.W. (Hrsg.)
 (2000). Die Diplomarbeit. Stuttgart: Schäffer-Poeschel

Prof. Dr. Heinz Lohse

Bewertung von Diplomarbeiten

1. Zur Bewertung von Diplomarbeiten

Die Betreuung und Bewertung von Diplomarbeiten gehört zu den Aufgaben des Hochschullehrers. Je nach der Stärke seines Engagements und Interesses bei der Themenpräzisierung und während des Entstehens der Diplomarbeit ergeben sich für ihn als Betreuer **unterschiedliche Grade des Bewertungsaufwandes**.

Bei der Bewertung von Diplomarbeiten steht die *Sache*, das ›Wie‹ der bearbeiteten Thematik, der Lösung der gestellten Aufgabe, der Erreichung des angestrebten Ziels im Vordergrund. Diese Sache ist von Wissenschaft zu Wissenschaft, ja von Wissenschaftsbereich zu Wissenschaftsbereich so unterschiedlich, daß es als vermessen gelten kann, wenn hier versucht wird, einige allgemeingültige Kriterien für die Urteilsfindung aufzustellen. Der Autor ist sich dieses Tatbestands durchaus bewußt, verfolgt auch keinesfalls die Absicht, den komplizierten und komplexen Vorgang der Bewertung einer Diplomarbeit zu normen oder gar zu reglementieren, nein – er tritt im Gegenteil dafür ein, die Eigenständigkeit und Eigenverantwortung des Hochschullehrers (Betreuers, Forschers) zu erhalten und zu stärken. Dennoch hält er es für nützlich, über die Bewertungsproblematik nachzudenken.

Schon die Vergabe der Themen von Diplomarbeiten hat vertrauensvoll, gewissenhaft und mit Blick auf das Wissen und Können des betreffenden Absolventen zu erfolgen. Dem Leistungsstärkeren wird man eine anspruchsvollere Aufgabe zumuten als dem weniger Leistungsfähigen, dem ›Tüftler‹ ein anderes Thema anvertrauen als dem mehr ›philosophisch Interessierten‹. Der Diplomand will nicht unterfordert, darf aber i.allg. auch nicht überfordert werden. Bei der Bemessung des Schwierigkeitsgrads der Aufgaben das richtige, auf hohe Leistungsmotivation beim Diplomanden hin orientierte Maß zu finden, gehört zu den wichtigsten Pflichten bei der Vergabe von Diplomarbeiten.

Grundbedingung für die Bewertung von Diplomarbeiten ist das Vorhandensein zweier Subjekte (od. Subjektgruppen):

- Gutachter oder Beurteiler,
- Autor(en) der Diplomarbeit bzw. Beurteilte(r)

sowie der als Objekt (meist in schriftlicher Form) vorgelegten Sachleistung des Diplomanden.

Obwohl beide Subjekte i.allg. bemüht sind, ihr Bestes zu geben, ist jede erbrachte Leistung, aber auch Leistungsbeurteilung mit Mängeln und Grenzen behaftet. Die Ursache dafür liegt in der Sache der behandelten Thematik, in der Person des Beurteilten, aber ebenso in der Person des Beurteilers.

Es ist davon auszugehen, daß die gerechte, vorurteilsfreie und objektive Bewertung einer Diplomarbeit schon davon abhängig ist, in welchem Maße der **Beurteiler** folgende **Voraussetzungen** erfüllt, die zwar jedem als selbstverständlich erscheinen, aber doch nicht immer alle gegeben sind (weil die Beurteilung der Arbeit an einen Assistenten delegiert wird oder aus was für Gründen auch immer):

1. Der Beurteiler ist körperlich und geistig fit und physisch imstande, die erbrachte Leistung verantwortungsbewußt zu bewerten. (Beispiel: Ein Univ.-professor, der 25 Diplomanden mit unterschiedlicher Arbeitsthematik gleichzeitig zu betreuen hat, ist nicht in der Lage, die 25 Arbeiten in einer begrenzten Zeit gründlich zu lesen und einzuschätzen.)
2. Der Beurteiler verfügt über die zur Bewertung der Arbeit erforderliche Sachkenntnis und das notwendige Sachverständnis und ist über den aktuellen Erkenntnisstand auf dem betreffende Gebiet bestmöglich informiert.
3. Der Beurteiler ist zur Erfassung und Bewertung der Arbeit motiviert. (Der Urteiler betrachtet die Einschätzung einer Diplomarbeit nicht als nebensächliche Aufgabe.)
4. Der Beurteiler ist von dem Willen durchdrungen, die vorgelegte Leistung objektiv, d.h. sachlich, exakt, gerecht und unabhängig von der betreffenden Person zu beurteilen. (Er erliegt nicht der Versuchung, die Arbeit aufgrund der Kenntnis der bisher erbrachten Leistungen des Diplomanden als zu gut oder zu schlecht ein-

zustufen.) Er läßt sich auch nicht von emotionalen Empfindungen gegenüber dem Autor beeinflussen.

5. Der Beurteiler besitzt die Fähigkeit zur richtigen und zweckentsprechenden Leistungseinschätzung. Diese Fähigkeit bezieht sich:
 - auf das Verständnis der Beziehung zwischen erwartetem Leistungsziel und vorgelegtem Ergebnis (Sachverständnis für die reale Möglichkeit des Vollzuges des Angestrebten), wozu auch Offensein für neue Gedankengänge und Lösungswege des Diplomanden gehört,
 - auf das psychologische und didaktisch-methodische Vermögen der vorurteilsfreien Bewertung der vorgelegten Arbeit,
 - auf das Vermögen der kritischen Beleuchtung und Bewertung des eigenen Erkenntnisstandes anhand der in der Arbeit dargebotenen Sachverhalte und
 - auf das Bestreben, weder dem ›Milde-Effekt‹ (der Betreuer urteilt i.allg. zu gut) noch dem ›Strenge-Effekt‹ (er urteilt zu hart) noch der ›Extremscheu‹ (er nutzt nur den mittleren Teil des Bewertungsmaßstabes) zu erliegen.

Empirische Untersuchungen – z.B. von TAFT (1962) – belegen, daß erhebliche Unterschiede bestehen in der Fähigkeit, wissenschaftliche Arbeiten zu beurteilen, Wert oder Unwert z.B. einer Diplomarbeit festzustellen.¹ Die Urteiler-Zuverlässigkeit ist gering (Reliabilitätskoeffizient $r < 0,56$, wobei $0 \leq r \leq 1$; zu den Gütekriterien siehe letzten Abschnitt). Dabei darf nicht unerwähnt bleiben, daß die Fähigkeit der Leistungseinschätzung abhängig ist vom Alter (Ältere urteilen i.allg. milder), vom Geschlecht, vom Erfahrungsschatz des Urteilers und seinem soziokulturellen Umfeld. Einen besonderen Einfluß auf die Bewertung von Arbeiten scheint das Selbstbild zu haben, das der Bewerter von sich bezüglich des bewerteten Sachverhalts hat.

¹ Nach TAFT zeichnen sich gute (Menschen-) Beurteiler u.a. durch folgende Merkmale aus:

- sachlich-nüchterne Einstellung (Physiker erzielten bei ihm bessere Ergebnisse als Psychologen),
- Selbstsicherheit, d.h. richtige Selbsteinschätzung, und
- emotionale Ausgeglichenheit.

2. Zur Methodik des Kriterienaufbaus

Normen und Werte sind die Grundlage jeder Art von »Bewertung«. Sollen **Kriterien für die Bewertung** von wissenschaftlichen Arbeiten herangezogen werden – und das ist häufig der einzig mögliche Weg –, hat man sich auf ein Normen- und Wertesystem zu beziehen. Das Problem besteht aber darin, daß sich Normen nicht rationell begründen, sondern eigentlich nur durch Verweis auf andere (übergeordnete oder grundlegendere) Normen rechtfertigen lassen. Leider besteht bislang kein inhaltlich bestimmtes Normungssystem, das allgemein akzeptiert ist und damit allgemein anwendbar wäre. Gibt es einen Ausweg aus diesem Dilemma?

Die zur Einstufung von Leistungen benötigten Kriterien fußen auf Wesensmerkmalen der Leistungsbeurteilung (kurz Merkmalen), die sich ihrerseits wieder in Kategorien bündeln lassen. Nach unserem Verständnis **liegen dem Bewertungsprozess also Merkmale zugrunde, die in Kategorien faßbar sind, aus denen sich die Kriterien formulieren lassen.** Die Merkmale beziehen sich

- auf die Beurteilung des Erfüllungsgrades der gestellten Aufgabe oder der Lösung des formulierten Problems,
- auf die Beurteilung von Selbständigkeit, Kreativität, Fähigkeiten zur Selbstkritik,
- auf die Beurteilung der Sprache und der äußeren Form der Arbeit usw., letztlich
- auf die Qualität der Arbeit als Ganzes.

An die Merkmale sind gewisse Anforderungen zu stellen, damit sie für die Leistungsbeurteilung auch wirklich brauchbar sind. Sie sollten (in Anlehnung an LATTMANN (1975))

- **erheblich** sein, d.h. wesentliche Kennzeichen der Leistungen erfassen,
- **allgemeingültig** sein, d.h. auf alle Diplomarbeiten anwendbar sein,
- **vollständig** sein, allen relevanten Kennzeichen einer Diplomarbeit Rechnung tragen,
- **unterscheidbar** sein, d.h. auseinandergelassen werden können,
- **voneinander unabhängig** sein, d.h. die Bedeutungskreise der einzelnen Merkmale sollten sich möglichst nicht überschneiden,

- **eindeutig** sein, d.h. von allen Beurteilern in semantisch gleicher Weise angewandt werden,
- **verstehbar** und **sinnvoll** und
- sie sollten **anwendbar** sein, d.h. zur Beurteilung der Diplomarbeit herangezogen werden können.

Zur Unabhängigkeit und zur Anzahl der verwendeten Merkmale lohnen sich ein paar weitere Überlegungen:

Es ist nicht zu leugnen, daß gewisse Merkmale sich bzgl. ihrer Semantik überlappen. Die Wahrscheinlichkeit dafür ist umso größer, je mehr Merkmale verwendet werden. Infolgedessen darf man es mit der Anzahl der Merkmale (und damit letztendlich Kriterien) nicht zu weit treiben. Beurteilungen aufgrund von 10 Merkmalen können zu den gleichen Bewertungen führen wie solche, denen 40 Merkmale zugrundegelegt werden.

Für eine relative Beschränkung der Zahl der Merkmale sprechen aber nicht nur statistische, sondern auch motivationspsychologische Überlegungen. Wenn der Beurteilung 40 Merkmale zugrundegelegt werden, so beträgt der durchschnittliche Einfluß des einzelnen Merkmals auf das Gesamtergebnis nur 2,5%. sind es dagegen 10 Merkmale, so schlägt jedes einzelne mit 10% zu Buche. Die Hauptschwierigkeit bei der Beschränkung der Zahl der Merkmale besteht darin, daß man die für die Bewertung der Diplomarbeit **wesentlichen** erkennen, herausfiltern muß. Sind es die relevanten und sind es echte Leistungsmerkmale, die entscheidend sind für die Beurteilung des Gesamtwerkes Diplomarbeit? Und: Kann das Merkmal erfaßt, d.h. operational umschrieben werden? Das sind die grundlegenden Fragen, die eigentlich gesondert für jede Wissenschaftsdisziplin beantwortet werden müßten. Für die Auslese der relevanten Merkmale böten sich Faktor- und Diskriminanzanalyse an.

Wie gehen wir nun vor, um trotz des weiter oben genannten Dilemmas zu den gewünschten Kriterien zu gelangen? Wir beschreiten folgenden Weg:

1. Aufschreiben vieler möglicher Bewertungsmerkmale. Hier geht es also um die Frage nach dem Bewertungsgegenstand im weitesten Sinne.

⇒ Was ist bewertbar?

2. Auswahl der für die Bewertung der Arbeit relevanten (relevant erscheinenden) Merkmale (Hauptmerkmale).

⇒ **Was gilt es (eigentlich) zu bewerten?**

3. Bündelung der Hauptmerkmale in unterscheidbare, etwa gleichgewichtige Kategorien

⇒ **Welche Bewertungskategorien?**

4. Formulierung eines Kriteriums je Kategorie so, daß die Hauptmerkmale bestimmend sind.

⇒ **Welche Bewertungskriterien?**

5. Anlegen eines Wertmaßstabs an jedes Kriterium.

⇒ **Was ist »gut«, was ist »weniger gut«?**

Eigentlich müßte ein solcher Wertmaßstab für jedes Hauptmerkmal angelegt werden. Das wäre zu aufwendig und würde die Angelegenheit nicht mehr handhabbar machen. Außerdem fehlt für viele Merkmale die Möglichkeit der Verwendung exakter quantitativer Bestimmungsverfahren. Schon das Anlegen eines Wertmaßstabs je Kriterium ist nicht einfach, denn zu verschieden sind die Auffassungen über das »Gewicht«, die »Gewichtigkeit« des einen oder anderen Kriteriums, denn sie unterscheiden sich ja nach Art und Aufbau voneinander. Wir nehmen eine Grobunterscheidung in »Kriterien der Wichtigkeitsstufe 1« und »Kriterien der Wichtigkeitsstufe 2« und – auch um extreme Gesamturteile auszuschließen – eine Gleichgewichtigkeit der Kriterien innerhalb der betr. Wichtigkeitsstufe vor.

6. Festlegung eines Wertes pro Kategorie (Kriterium).

⇒ **Welcher (Punkt-)Wert ergibt sich je Kriterium?**

Wir entscheiden uns hier für eine Erfassung nach einer Intervallskala, die

für die Wichtigkeitsstufe 1 zwischen +2 $\hat{=}$ »voll bejahend« und -1 $\hat{=}$ »verneinend« vier Stufen zuläßt,

für die Wichtigkeitsstufe 2 zwischen +1 $\hat{=}$ »bejahend« und -1 $\hat{=}$ »verneinend« nur drei.

7. Verdichtung zu einem Gesamturteil.

⇒ **Welche Note ergibt sich für das Ganze?**

Es erhebt sich die Frage, ob sich ein allzu großer Aufwand wirklich lohnt und ob man nicht mit geringeren Mitteln die gleichen Resultate einer den Gütekriterien weitgehend genügenden Bewertung von Diplomarbeiten erzielt.

Entscheidend ist, in welchem Maße sich ein solches System in der täglichen Praxis bewährt.

Für drei (von 18) Bewertungskriterien sei das Vorgehen gemäß 1. bis 7. aufgezeigt:

Hauptmerkmal:	Bewertungskategorie:
»Erfassen des Themas« »Klarheit der Aufgaben-, Frage- bzw. Zielstellung«	1. Erfassen der Diplomaufgabe
»Konsequenz der Bearbeitung des Themas« »Prägnanz der Darstellung«	2. Art der Bearbeitung
»Eindeutigkeit der Formulierung von Definitionen, Sätzen, Urteilen usw.« »Gültigkeit von Argumentationsketten, Beweisen« »Logik der Gedankenführung«	3. Eindeutigkeit und Logik

Unter Beachtung der Dominanz der Hauptmerkmale wird je Kategorie ein Kriterium formuliert. Die hier entstehenden sind im nächsten Teilabschnitt (Kriterien der Wichtigkeitsstufe 1) unter 1., 2. und 3. zu finden. Es erwies sich – wie schon erwähnt – als zweckmäßig und für die objektive Gesamtbeurteilung als nützlich, sowohl bei der Auswahl der Hauptmerkmale als auch der Kriterien zwischen »Kriterien der Wichtigkeitsstufe 1« = »entscheidenden« und »Kriterien der Wichtigkeitsstufe 2« = »zweitrangigen« zu unterscheiden.

Der so entstandene Kriterienkatalog sei nun im weiteren vorgestellt.

3. Die Kriterien

Bei jeder Durchsicht sollten die folgenden allgemeingültigen Kriterien genutzt werden, aber keinesfalls in dogmatischen Sinne. So kann eine geniale Arbeit vorliegen, bei der aber einige Kriterien nur unzureichend erfüllt sind. Trotzdem ist die dann entsprechend ihrer vielleicht epochalen Bedeutung sehr positiv einzuschätzen.

Kriterien der Wichtigkeitsstufe 1 (Entscheidende Kriterien)

1. Ist das **Thema** der Arbeit **erfaßt** worden und gelang die Umsetzung des Themas in eine klare, eindeutige, begrifflich sauber **umrissene Aufgaben-, Frage- bzw. Zielsetzung**?
2. Wird diese Aufgaben-, Frage- bzw. Zielstellung während der Abhandlung der Arbeit **konsequent verfolgt**, sind also keine vom Thema abweichenden oder zu breit ausgeführten Passagen enthalten?
3. Sind Definitionen, Prämissen, Lehrsätze, Behauptungen, Urteile zweifelsfrei formuliert und problemstellungsgemäß durchgehalten, und sind Argumentations- und Belegketten bzw. Beweise lückenlos vorhanden und in sich widerspruchsfrei, ist die **Gedankenführung logisch**?
4. Entspricht die Arbeit als Ganzes dem jetzigen Erkenntnisstand der entsprechenden Wissenschaft, hat sie das zu erwartende Ni-

veau und die erforderliche Tiefe und bringt sie einen **nachweisbaren Gewinn für Theorie oder Praxis**?

5. Entsprechen die angewandten **Untersuchungs- (Forschungs-) methoden dem heutigen Stand** des entsprechenden Wissenschaftszweiges und werden sie konsequent und fehlerfrei angewandt?
6. Erfolgt die **Auswertung** der Untersuchungen **wissenschaftlich einwandfrei**?
7. Ist die **Ergebnisdarstellung klar**, eindeutig und exakt, entsprechen die Ergebnisse den am Beginn der Arbeit formulierten Aufgaben-, Frage- bzw. Zielstellungen und erfüllen sie weitgehend die an die Umsetzung billigerweise zu stellenden Erwartungen?
8. Wurde zur Bewältigung des Themas qualitativ angemessene **Literatur** herangezogen und wurde die im Quellenverzeichnis angeführte **Literatur tatsächlich und in korrekter Weise** (z.B. Zitierweise, Art und Umfang der Zitate betreffend) **verwendet**?
9. Zeigt die Arbeit **Selbständigkeit**, Eigenständigkeit, Originalität, Kreativität, Mut zu neuen Gedankengängen, setzt sie sich kritisch mit bisherigen Auffassungen, Erkenntnissen, Lehrmeinungen auseinander?
10. Entsprechen **Sprachstil und Wortwahl** den Erfordernissen einer wissenschaftlichen Qualifikationsarbeit? (Richtiger Einsatz wissenschaftlicher Termini, Verständlichkeit und Prägnanz des Textes, Logik der Wortwahl und Satzfolge etc.)

Bei empirischen und experimentellen Untersuchungen treten folgende Hauptkriterien zusätzlich auf:

11. Wurden **exakt formulierte** Forschungs- (und gegebenenfalls statistische) Hypothesen aufgestellt und diese **fach- und sachgerecht geprüft**?
12. Existieren saubere **Versuchs- und Variablenpläne**, entsprechen die Untersuchungsmethoden und Auswertungstechniken den Anforderungen moderner Wissenschaftstechnologie (insbes. der

Computertechnik) und ist das Auswertungsdesign klar dargestellt, vollständig offengelegt und – z.B. anhand der Anlagen der Arbeit – nachvollziehbar?

Kriterien der Wichtigkeitsstufe 2 (zweitrangige oder Nebenkriterien)

- N1. Ist die **Gliederung formal korrekt und logisch schlüssig**?
- N2. Werden – vom wissenschaftlichen Niveau her – **keine Selbstverständlichkeiten** oder Trivialitäten formuliert und Wiederholungen vermieden?
- N3. Ist die **äußere Form** der Arbeit (einschließlich Titel- und Seitengestaltung) den Vorschriften entsprechend und ansprechend?
- N4. Folgt das Geschriebene den Regeln der **deutschen Rechtschreibung**, das (evtl.) Dargestellte den DIN-Bedingungen, ist die eidesstattliche Erklärung vorhanden und unterschrieben?
- N5. Sind **Abbildungen und Tabellen** exakt durchnummeriert und inhaltlich bezeichnet, sind die erforderlichen Verzeichnisse (Inhalts-, Abbildungs-, Abkürzungs-, Literatur- und Quellenverzeichnis) vorhanden und qualitativ einwandfrei?
- N6. Ist der **Umfang der Arbeit** der gestellten Aufgabe angemessen?

Nun kann man auf folgende Art und Weise zu einer formalen Bewertung der Arbeit gelangen:

- Kriterien der Wichtigkeitsstufe 1:

Kann die jeweilige Frage voll bejahend beantwortet werden (**die Anforderung wird ausgezeichnet oder sehr gut erfüllt**), so erteilt man darauf 2 Punkte, kann sie **bejahend (gut erfüllt)** beantwortet werden, so gibt man einen Punkt; wird sie verneint (**die Anforderung wird nicht erfüllt**), gibt es einen Minuspunkt. Kann man sie **weder bejahen noch verneinen**, fällt sie raus (**0 Punkte**).

- Kriterien der Wichtigkeitsstufe 2:

Kann die jeweilige Frage im großen und ganzen **bejahend** beantwortet werden, so erteilt man darauf **einen Punkt**, wird sie **verneint**, gibt es **einen Minuspunkt**. Kann man sie **weder bejahen noch verneinen**, fällt sie heraus (**0 Punkte**).

Die erzielten Punktwerte sind zu addieren.

Auf diese Weise ergibt sich für die Mehrzahl der Diplomarbeiten eine **Höchstpunktzahl von 26 Punkten** und eine **Mindestpunktzahl von -16 Punkten**, handelt es sich um eine Diplomarbeit mit **empirischer** Untersuchung, dann gilt **Höchstpunktzahl: 30 Punkte** und **Mindestpunktzahl -18 Punkte**.

Für diese beiden Bereiche »26... – 16 Punkte« bzw. »30... – 18 Punkte« sind dann abschnittsweise Gesamurteile festzulegen. So könnte für das Intervall »26 u. 25 Punkte« (bei empirischen Arbeiten »30 u. 29 Punkte«) das Prädikat **»Ausgezeichnet«** ausgesprochen werden, für »24 bis 21 Punkte« (bzw. »28 bis 24 Punkte«) das Prädikat **»Sehr gut«**. Die Festlegung der entsprechenden Intervallgrenzen bleibt aber weitgehend der Eigenverantwortung des jeweiligen Beurteilers überlassen, denn er kennt die gestellten Anforderungen und das Bewertungsumfeld am besten, orientiert sich dabei jedoch an den Diplomprüfungsordnungen seiner Einrichtung.²

2 Es gehört zu den grundsätzlichen Empfehlungen an einen Diplomanden, die jeweilige Prüfungsordnung bereits vor Beginn der Arbeit intensiv zu studieren. Mit dieser Kenntnis und dem oben veröffentlichten Bewertungsschema lassen sich sehr gute Hinweise auf die mögliche Bewertung der Arbeit ableiten. Um noch Korrekturen auch im Falle grober Abweichung zu ermöglichen und evtl. Defizite aufzuzeigen, empfiehlt es sich, das Bewertungsschema unter Berücksichtigung des Standes der Arbeit bereits das erste Mal zu Beginn des letzten Drittels der Arbeitszeit zu Rate zu ziehen und eine »Probewertung« vorzunehmen. (Anmerkung des Herausgebers)

4. Bemerkung zu Problemen der Bewertungsqualität

In nicht wenigen Wissenschaftsdisziplinen ist immer wieder festzustellen, daß zwei oder mehrere Beurteiler von Diplomarbeiten zu ganz unterschiedlichen Einschätzungen des Wertes der Arbeit kommen. Das kann viele Gründe haben, die in der Person des Beurteilers liegen (Unterschiede im Alter, im Anforderungsniveau, im Erkenntnisstand; Mangel an Stabilität der zu betrachtenden Aspekte des Bewertungsvorgangs usw.).

Die Hauptsache für diese Urteils-Variabilität liegt aber darin, daß den angewandten Bewertungsverfahren keine **Methodeneichung** zugrundegelegt wird. Die Qualität einer Bewertungsmethode hängt – wie bei jeder Methode – wesentlich davon ab, inwieweit eine Eichung des Verfahrens vorgenommen wurde, d.h. inwieweit die Haupt-Gütekriterien einer Prüfung unterzogen worden sind. Die Gütekriterien sind methodentheoretisch notwendige und hinreichende Indikatoren zur Abschätzung der Leistungseigenschaften der betreffenden Bewertungsmethode. Unter Methodeneichung versteht man also die Bestimmungen der Haupt-Gütekriterien, als da sind Konkordanz (früher oft als Objektivität bezeichnet), Reliabilität und Validität. Näheres zu diesen Gütekriterien siehe in Abschnitt 4 des Beitrages »Empirische Untersuchungen«

Auch die vom Autor oben vorgeschlagene Bewertungsmethode müßte der Methodeneichung unterzogen werden. Es versteht sich von selbst, daß das ein sehr aufwendiges Verfahren ist, es ja für Hunderte von Wissenschaftsdisziplinen, in denen Diplomarbeiten geschrieben werden, getrennt erfolgen müßte.

Als offenes Problem bleibt die **Frage, ob eine derartige Vereinheitlichung der Bewertung tatsächlich anzustreben ist**. Auch auf die Gefahr von Fehlbeurteilungen hin ist es in Anbetracht der Komplexität des menschlichen Geistes und der Vielfalt und Vielschichtigkeit des von ihm Untersuchten, damit in Anbetracht der Komplexität des Bearbeitungsprozesses eines Themas und – nicht zuletzt – der Bewertung der resultierenden Arbeit vernünftigerweise gar nicht anzustreben, den Bewertungsvorgang zu normieren.

5. Schlußbetrachtung

Es sei zum Schluß ausdrücklich darauf verwiesen, daß bei allen Bewertungen, die von einem Beurteiler vorgenommen werden, die einschlägigen Gesetze, Rechtsvorschriften, Prüfungsordnungen etc. des jeweiligen Landes oder/und der Universität (Hochschule), an der die Diplomarbeit eingerichtet wird, einzuhalten sind.

Dabei steht das Grundgesetz – besonders hinsichtlich der Unantastbarkeit der Würde des Menschen (Artikel 1 GG) – obenan.

Literaturverzeichnis

- Brandstetter, Hermann: Die Beurteilung von Mitarbeitern. In: Handbuch der Psychologie, 9. Band, Göttingen 1970
- Clauß, Günther u.a. (Hrsg.): Wörterbuch der Psychologie, Bibliographisches Institut, Leipzig 1985
- Lattmann, Charles: Leistungsbeurteilung als Führungsmittel. UTP, Berlin und Stuttgart 1975
- Schaerli, Anton: Praxis der Leistungsbewertung. In: Industrielle Organisation. Hrsg.: Betriebswissenschaftl. Institut der Eidgenöss. Technischen Hochschule Zürich. 11 (37. Jg.) 1968
- Sprung, Lothar/Sprung, Helga: Grundlagen der Methodologie und Methodik der Psychologie, Berlin 1984
- Strametz, Dieter/Lometsch, Arndt: Leistungsbeurteilung in deutschen Unternehmen, Königstein/Ts. 1977
- Traft, Ronald: The ability to judge people. Performance appraisal. Hrsg.: Th.L. Whisler and Sh. F. Harper, New York 1962
- Theisen, René: Wissenschaftliches Arbeiten, Technik – Methodik – Form, München 1993
- Zander, Ernst: Arbeits- und Leistungsbewertung, Heidelberg 1970

wäre, als mit meinem Kriterien-Katalog neben sich die Examensarbeit zu schreiben. Oder mit anderen Worten: meine Anforderungen an eine Examensarbeit bzw. an eine/n Studierenden bestehen nicht darin, daß sie oder er die Arbeit nach meinen Vorstellungen, Anforderungen, Kriterien schreibt, mir also „nach dem Munde“ schreibt. Wer das täte, begäbe sich der Chance, seine eigene, selbstbestimmte Arbeit zu schreiben. Vielmehr geht es darum, gemeinsam einen Minimal-Consens zu finden, der für beide Seiten befriedigend ist. Und weiter geht es darum, daß Autor/in einer Examensarbeit (auch) eigene oder weitere Kriterien entwickelt als die meinen, wenn sie/er davon überzeugt ist, das tun zu müssen. Wenn sie sich mit den meinen treffen, so habe ich nichts dagegen. Wenn sie andere sind als die meinen, so werde ich sie in ihrer Ernsthaftigkeit prüfen und dann akzeptieren, gegebenenfalls aber verwerfen (müssen). Besonders das Eintreten dieses Falles wird gründlich zu besprechen sein, möglichst während des Entstehens der Arbeit, nicht erst nach ihrer Fertigstellung. Das aber hängt von der Bereitschaft der ExaminandInnen ab; ich bin dazu bereit, halte das sogar für eine meiner Berufspflichten. – Der Katalog kann bei weiteren Studierenden erneut Angst auslösen. Auch da hilft nur geduldiges gemeinsames Bearbeiten dieser Angst. Denn nichts ist ein für allemal bearbeitet.

Kriterien-Katalog zur Abfassung und Beurteilung von Examensarbeiten im Hauptfach Erziehungswissenschaft

(MA, Diplom, Dissertation)

Vorbemerkungen:

Dieser Katalog gibt alle möglichen Kriterien wieder; „alle möglichen“ ist nicht mißzuverstehen im Sinne eines willkürlichen Sammelsuriums, sondern in diesem Sinne: das sind die

Kriterien, die ich sehe, mehr habe ich nicht, ich habe also auch keine in der Hinterhand. – Das heißt aber nicht, daß alle Kriterien auf alle Arbeiten angewendet werden (müssen oder können). Siehe dazu auch die „Ergänzungen“.

Prinzipiell gibt es für mich keinen Unterschied zwischen Diplom-, Magister- und Doktorarbeit, denn bei allen handelt es sich um wissenschaftliche Arbeiten. Bezeichnenderweise rangiert die Dissertation im öffentlichen Bewußtsein wie im universitären Selbstverständnis aber an erster Stelle, gilt also offensichtlich als „noch wissenschaftlicher“. Möglicher Unterschied zwischen Diplom- und Magisterarbeit auf der einen Seite und Dissertation auf der anderen Seite in meinen Augen: Dissertationen sollten, schon weil dafür mehr Zeit zur Verfügung steht, wohl die Kriterien 1.10, 1.11 (quantitativ wie qualitativ) erfüllen, gegebenenfalls auch 1.13, 1.14.

1. *Inhaltliche Kriterien:*

- 1.1 möglichst genaue Angabe der Motivation zur Arbeit bzw. zur Thematik (subjektiv / transsubjektiv); wenn vorhanden: Verdeutlichung der subjektiven Betroffenheit, aber nicht nur innerhalb der Motivation, sondern durchgehend; Begründung der Themenwahl durch Nachweis der theoretischen und/oder praktischen Relevanz;
- 1.2 genaue Angabe des (Untersuchungs-)Gegenstandes, der Fragestellung(en) und/oder des Ziels der Untersuchung (das braucht nicht immer nur zu Beginn der Arbeit der Fall zu sein, sondern kann entwickelt, d. h. laufend unternommen werden);
- 1.3 Bestimmung (Entscheidung für) und dann in diesem Sinne konsequenter Gebrauch verwendeter Fachausdrücke / Begriffe;
- 1.4 möglichst genaue Beschreibung und einleuchtende Be-

- gründung der gewählten / verwendeten Methoden (einschließlich des Verfahrens der Auswahl und Analyse von Fachliteratur);
- 1.5 konsequente und zuverlässige Anwendung der Methoden (auch: Repräsentativität oder begründete Nichtrepräsentativität der ausgewählten Literatur und Qualität ihrer Rezeption und Kritik);
 - 1.6 klare Gedankenführung (Verdeutlichung des jeweiligen und des übergeordneten, größeren Begründungszusammenhangs); Differenziertheit des Urteils und Einschätzungsvermögens;
 - 1.7 Berücksichtigung verschiedener Möglichkeiten der Interpretation und Erklärung von Ergebnissen (fremden und eigenen Ergebnissen);
 - 1.8 selbstkritische Haltung auf der einen Seite, Position beziehen auf der anderen Seite (das ist kein Sowohl-als-auch, sondern eine dialektische Beziehung);
 - 1.9 Erreichen des Untersuchungszieles, d.h.: wird der Anspruch verwirklicht, den die Arbeit (und damit ihre SchreiberIn) an sich selber stellt?
 - 1.10 Originalität, Kreativität (früher: „selbständiger Beitrag zur Forschung“ bzw. „Schließen einer Forschungslücke“); vgl. auch unter 1.11 bis 1.16;
 - 1.11 Schwierigkeitsgrad der Aufgabe(n); vgl. auch unter 1.10 bis 1.16;
 - 1.12 um was für eine Arbeit handelt es sich? (Typen: literarische (theoretische) Arbeit, empirische Untersuchung, Handlungsforschungsansatz, [Selbst-]Erfahrungsbericht, Praxisbericht; Mischtypen). Keine Unterschätzung von Erfahrungsberichten und den mit ihnen verbundenen Möglichkeiten und Schwierigkeiten! „Pionier“-Arbeiten ebenfalls erwünscht, d.h. z.B. solche,

für deren Gebiet / Thema bisher wenig Fachliteratur vorliegt. Das gilt insbesondere für autobiografische / selbstreflexive Arbeiten. Aber Vorsicht: sie sind weitaus schwieriger als übliche wissenschaftliche Arbeiten, können also auch eher „danebengehen“;

- 1.13 Fähigkeit (und Möglichkeit), Theorie und Praxis miteinander zu verbinden bzw. einander anzunähern (oder auch z.B. Empirie und Hermeneutik), d.h.: kein einseitiges, abstrakt abgehobenes Theoretisieren und: kein einseitiges, blindes Praktizieren im Sinne von z.B. Rezeptologie;
- 1.14 Fähigkeit (und Möglichkeit) interdisziplinärer Kooperation;
- 1.15 gesellschaftliche/r Relevanz / Bezug der bearbeiteten erziehungswissenschaftlichen Thematik (kein monokausaler Begründungszusammenhang! Monokausal wäre es z.B., alle Schuld oder Verantwortung entweder nur der Gesellschaft oder nur dem Individuum zuzuschreiben);
- 1.16 Aufzeigen / Herausarbeiten von realistischen (oder auch utopischen) Perspektiven / Alternativen.

2. Formale Kriterien:

- 2.1 differenziertes Inhaltsverzeichnis (am besten nach Dezimalsystem), aber nicht überladen! (d.h.: Unter-Unterpunkte brauchen nicht unbedingt im Inhaltsverzeichnis aufzutauchen, können im laufenden Text als Kernwörter hervorgehoben, unterstrichen werden);
- 2.2 „aufschließende“ Gestaltung und Gliederung des laufenden Textes; Stichwort: „Leserfreundlichkeit“ (z.B. Hervorheben von Kernwörtern, Begriffen, Aspekten u.ä. durch Unterstreichen; Einrücken und/oder Eng-rücken von Zitaten);

- 2.3 Vorsicht: nicht zu viele Zitate! (Gefahr zu vieler verschiedener Stile und Verlust der eigenen Sprache); Zitate nicht als Autoritäten auffahren; nur dann, wenn ein Autor es „unübertrefflich“ ausdrückt, sonst Inhalt mit eigenen Worten wiedergeben; Zitate können aber auch veranschaulichen;
- 2.4 übersichtliche Zusammenfassung von Hauptteilen, Hauptkapiteln, Ergebnissen, Zwischenergebnissen;
- 2.5 Ermöglichung der intersubjektiven Nachprüfung; das gilt nicht nur für empirische Arbeiten (Materialien, Tabellen, Schemata, Fragebögen, Interview-Leitfaden u. ä.; Anhänge, Anhangsband);
- 2.6 klare, prägnante Sprache, ohne Bluff und akademischen Qualm, Vermeidung des unpersönlichen Wissenschaftsjargons („man“, „wir“, „der Verfasser“, „es ist festzustellen“ usw.); eigene Sprachfindung, jeweils dem Gegenstand angemessen, z.B. ganz konkret, bildhaft, ironisch u. ä.;
- 2.7 Auslagerung weiterführender Einzelheiten, Nebenaspekte, Exkurse, statistischer Daten, weiterführender Literaturhinweise u. ä. in die Anmerkungen (besser, als Anmerkungen ans Ende des Gesamttextes zu setzen, ist es, sie unten auf die laufenden Seiten zu placieren – leserfreundlicher!);
- 2.8 Berücksichtigung geltender Regeln für die formale Abfassung wissenschaftlicher Arbeiten: einwandfreie Zitieren (Kennlichmachung), Sorgfalt bei der Bezifferung von Anmerkungen (keine Anmerkung überspringen u. ä.); korrekte Seitenverweise (oder lieber gar keine), Korrekturlesen des fertigen Textes. Bei letzterem handelt es sich nicht unbedingt um Pedanterie, sondern um praktische Solidarität mit dem Leser, der ja nicht man selber ist. Dies gilt besonders für die Eindeutigkeit

und Unmißverständlichkeit der Sprache bzw. Ausdrucksweise.

Ergänzungen zum Kriterien-Katalog:

1. Eine Examensarbeit ist eine geschriebene Arbeit („geschrieben“ nicht mit „schriftlich“ zu verwechseln, obwohl beides miteinander zu tun hat. „Geschrieben“ soll hier vor allem heißen „abgeschlossen“, „fertig“, „beendet“, und das ist insofern absurd, als die damit verbundenen Lernprozesse lebenslang nicht abschließbar sind). Zum Schreiben selbst: Schreiben und dann auch Lesen der Arbeit sollen nicht nur Arbeit, sondern nach Möglichkeit auch Freude machen, faszinieren, eigene Lernprozesse initiieren. Stichwörter: Selbstbetroffenheit, Selbstbestimmung, Kreativität. Aber: „Das Werk ist die Totenmaske der Konzeption“ (W. Benjamin: Einbahnstraße. Frankfurt/M., 1972, S. 43). Ist die Freude aber auch immer der Tod der Vorfreude? – Der Kriterien-Katalog macht u.U. deshalb Angst, weil er (auch) diese „Totenmaske“ aufzeigt (Literatur zum Schreiben: Barthes, R.: Die Lust am Text. Frankfurt/M., 1974; Kerouac, J.: Unterwegs. Reinbek, 1968; der oben erwähnte Text von Benjamin).
2. Der Kriterien-Katalog stellt durchaus so etwas wie eine Entwicklungsfolge dar, vor allem seine inhaltlichen Kriterien (d.h. zeitliche Rangfolge, z.B. von „Motivation“ [Anfang] bis „Aufzeigen von Perspektiven / Alternativen“). Trotzdem erhalte ich meine den Katalog einleitende Aussage aufrecht, nämlich: nicht alle Kriterien (müssen oder können) auf alle Arbeiten angewendet (werden). Welche das sind oder sein können, muß oder kann mit dem/der konkreten Studierenden für seine/ihre konkrete Arbeit besprochen werden.
3. Unterschiedliche Kategorien / Typen von Arbeiten (Katalog, Punkt 1.12: literarische, empirische Arbeiten usw.):

ich gruppiere folgendermaßen:

- a) herkömmliche wissenschaftliche Arbeiten (Literaturarbeit, empirische Untersuchung): ich bezeichne sie als „abständige“, abstrakt-abgehobene Arbeiten, in die das schreibende Subjekt sich möglichst wenig einbringen soll (für diese Arbeiten treffen vor allem zu die Kriterien 1.2 bis 1.5, 1.7, 1.14, 2.5);
- b) neue wissenschaftliche Arbeiten (Handlungsforschungsansatz, Erfahrungsbericht (eigene Praxis, Selbsterfahrung), Mischtypen (z.B. Konfrontation Erfahrung / Literatur), autobiographische / selbstreflexive Arbeit): ich bezeichne sie als Arbeiten, in die das schreibende Subjekt stark eingeht. Auch hier treffen die unter a) genannten Kriterien zu, aber quasi „in neuer Qualität“; weiterhin besonders die Kriterien 1.1, 1.6, 1.8 bis 1.11, 1.13, 1.15, 2.3, 2.6.

Selbsterfahrungsbezogene / autobiografische / selbstreflexive Arbeiten müssen nach meinen bisherigen Erfahrungen damit besonders folgendes beachten:

- schonungslose Offenheit / Aufrichtigkeit sich selber gegenüber;
- daraus folgt: Stehenlassen (nicht: Wegargumentieren) von Widersprüchen, Ambivalenzen u. ä.;
- die Stufe des Narrativen (Erzählenden) darf keineswegs -zig Seiten umfassen, sondern sollte „auf den Begriff gebracht werden“ (d.h. Herausarbeitung von Trends, Richtungen, „Mustern“);
- Begrifflichkeit und Sprache müssen hier besonders „scharf“, deutlich, eindeutig, unmißverständlich sein, um den Leser, bei aller vorausgesetzten Empathie, verstehen zu lassen;
- „Verallgemeinerungsfähigkeit“ sollte nicht übers Knie ge-

brochen werden. Erprobte Hilfsmittel sind: Konfrontation mit ausgewählter, themenbezogener Fachliteratur; ein Gesprächspartner („Führer“, Supervisor), der dann durchaus im Text auftauchen kann mit seinen Aussagen, Stellungnahmen u. ä.; weitere (Kombination von) Methoden, z.B. Interviews, Fallstudien, Protokolle u. ä.;

- „Verallgemeinerungsfähigkeit“ muß aber nicht erreicht werden, ebensowenig ist Literatur-Einbeziehung um jeden Preis zu empfehlen, vor allem dann nicht, wenn mit der Literatur halbherzig umgegangen wird, im Sinne einer lästigen Pflichtübung. – Das erfordert dann aber hohe „Einzelqualität“; siehe dazu alle Punkte vor „Verallgemeinerungsfähigkeit“;
- hinsichtlich der Signalisierung von „Verallgemeinerungsfähigkeit“ bzw. „Nichtverallgemeinerungsfähigkeit“: Vorsicht bereits bei der Formulierung des Themas der Arbeit! Beispiel: wenn eine einzige Frau von sich selber schreiben will, sollte sie nicht von „Frauen“ sprechen. – Thema kann mit Neben- oder Untertitel versehen werden. Thema nicht zu eng fassen und auch mit „Wegweiser“-Funktion versehen! (Für künftige Leser wichtig). Versuchen, sich selber in der Themenformulierung wiederzufinden bzw. sich mit dem Thema identifizieren zu können; das aber nicht bis hin zur Selbstaufgabe oder zur Fixierung auf das Thema. – Auch diese Ergänzungen werden weiterhin zu ergänzen sein.

Der Katalog ist in der Tat sehr knapp formuliert und bedarf trotz seiner vorhandenen Unmißverständlichkeit immer wieder der zusätzlichen mündlichen Erörterung mit den Studierenden. Sinn bzw. *Absicht eines Kataloges ist es, in Kürze zu informieren*, und das tut mein Katalog. Das mußte ich auch einem Kreis von Kollegen sagen, denen ich ihn vor einigen Jahren vorstellte und deren Haupteinwand auf seine Kürze bzw. Knappheit zielte. Es ist bezeichnend, daß dagegen eine Kolle-