

LUTZ MÜLLER & DIRK BÜSCH

Peer-Review in der Sportwissenschaft – Erfahrungen mit und Konsequenzen aus dem Begutachtungsverfahren beim 3. Sportspiel-Symposium der dvs 2002 in Bremen

1 Einleitung

Im Rahmen des 3. Sportspiel-Symposiums der dvs „Begründungsdiskurs und Evaluation in den Sportspielen“ im September 2002 in der Universität Bremen haben wir innerhalb der Kommission ein transparentes und verbindliches Begutachtungsverfahren („peer-review“) der eingereichten Beiträge unternommen. Die Intention einer wissenschaftlichen Qualitätssicherung verband sich so selbstbezüglich mit der Thematik der Tagung und ihren Beiträgen. Im Zusammenhang der dabei aufgetretenen Probleme und Erfahrungen möchten wir eine zweite Ebene der Selbstbezüglichkeit zur Diskussion stellen, indem wir die Thematik „Begründungsdiskurs und Evaluation“ nunmehr auf das Begutachtungsverfahren selbst beziehen.

Dazu werden wir eingangs kurz die Intentionen und Verfahrensregeln des zum Sportspiel-Symposium angewandten Verfahrens umreißen. Daran schließt sich eine systematisierte Sicht auf einschlägige Problemlagen an, die im Laufe des Verfahrens und reflexiv deutlich geworden sind. Unsere Erfahrungen und deren Analyse münden in einen Vorschlag für künftige Verfahrensweisen. Dieser Vorschlag wird der sportwissenschaftlichen Scientific Community zur Diskussion gestellt.

2 Zum Begutachtungsverfahren für das 3. Sportspiel-Symposium

Das örtliche Organisationskomitee verband mit seinen Planungen für das Sportspiel-Symposium die Intention, zu einem möglichst hohen wissenschaftlichen Standard dieser Veranstaltung beizutragen. Dazu wurde ein Maßnahmenkatalog entwickelt, der zur Strukturierung der Planungsphase diente:

- Von einem ‚ex post‘ gefertigten Berichtsband wurde zugunsten eines vorab publizierten Abstractbandes Abstand genommen, u.a. um der langen Produktionsdauer und den bekannten inhaltlichen Disparitäten solcher Berichtsbände zu entgegenen.
- Die mit Fristsetzung eingereichten Abstracts sollten entsprechend formalen Vorgaben (Manuskriptgestaltung gemäß APA bzw. neuem dvs-Standard) gestaltet werden. Bei Nichterfüllung der Vorgaben wurden die Abstracts mit einem Hinweis der formalen Überarbeitung zurückgesandt. Um das Verfahren zu vereinfachen, wurden auf der Homepage des Symposiums eine Dokumentvorlage zur Verfügung gestellt und partiell konkrete Überarbeitungshinweise gegeben.
- Das ‚peer-review‘-Verfahren erfolgte als ‚Doppel-Blind-Verfahren‘, d.h. anonymisiert durch die Mitglieder des wissenschaftlichen Beirats. Die Kriterien der Begutachtung wurden zur Orientierung in Form eines Evaluationsbogen auf der Homepage frühzeitig veröffentlicht. Die jeweiligen Ergebnisse der Gutachten sollten zu Entscheidungen über die Annahme,

zu Hinweisen auf erforderliche Überarbeitungen und ggf. auch zur begründeten Ablehnung von Beiträgen führen. In dem Fall, dass beide Gutachten für eine Ablehnung plädierten, wurde vom Organisationskomitee ein drittes, jedoch kein vergleichendes Gutachten in Auftrag gegeben, d.h., dem Gutachtenden war nicht bekannt, dass es sich um ein Drittgutachten handelte. Das gegenüber Zeitschriften moderate Begutachtungsverfahren sollte unter anderem auch den Besonderheiten, z.B. nur wenige eingereichte Abstracts usw. gerecht werden.

- Mit der Rückmeldung der Begutachtungsergebnisse an die Autoren/innen wurde eine hinreichende, d.h. 14-tägige Überarbeitungsfrist eingeräumt. Sofern dann immer noch erhebliche Differenzen zu den gutachterlichen Hinweisen bestanden, wurde eine erneute Rückkoppelung zwischen dem Organisationskomitee und den Betroffenen durchgeführt.
- Die teilweise mehrmals und dann beim Organisationskomitee nochmals von (formalen) Mängeln bereinigten Abstracts wurden dann in einem Abstractband zusammengefasst und nach dem Druck ca. einen Monat vor Beginn des Symposiums allen angemeldeten Teilnehmenden zugesandt.

Zusammenfassend lassen sich als grundlegende (intendierte) Prinzipien dieses Verfahrens hervorheben: (1) Qualitätssicherung im Hinblick auf definierte formale und inhaltliche wissenschaftliche Standards, (2) ein Höchstmaß an kriterienbezogener wie organisatorischer Transparenz bei gleichzeitig strikter Vertraulichkeit und (3) schließlich eine hohe terminliche Verbindlichkeit.

3 Probleme und Irritationen

Unsere Erfahrungen mit diesem Review-Verfahren – die wir als Organisationskomitee und nicht als Referenten berichten – lassen sich in einer systematisierten Problemübersicht beschreiben. Dabei werden insbesondere auch Unterschiede in Handlungsweisen und Reaktionen von Beteiligten verdeutlicht.

- Generell wurde die Intention einer wissenschaftlichen Qualitätssicherung nicht in Frage gestellt, vielfach sogar explizit als sinnvoll und notwendig eingeschätzt. Problematisch wurde es des Öfteren dann, wenn in den Gutachten ‚negative‘ Einschätzungen von Form und/oder Inhalt der Abstracts expliziert wurden. Die Bandbreite der Reaktionen reichte dann von Auffassungen einer Selbstevidenz der eigenen wissenschaftlichen Qualität über Unverständnis („Abstracts habe ich immer schon so und ohne Beanstandungen geschrieben“) bis hin zu einem tendenziellen Ignorieren formaler wie inhaltlicher Überarbeitungsempfehlungen. Verbunden war dies teilweise auch mit einem in der Tonart verschärften Schriftwechsel, zahlreichen Telefonaten

und vielen anderen unangenehmen Begleiterscheinungen. Dies führte u.a. zu einem organisatorisch vorab nicht hinreichend kalkulierten Arbeitsaufwand für alle Mitarbeiter/innen. Andererseits gab es (überwiegend) auch ein einverständiges Überarbeiten der Abstracts oder aber deren Zurücknahme auf Grund einsichtiger Anerkennung der Begutachtungen.

- Als ein leichter zu lösendes Problem erweist sich (heute) für uns der Umgang mit Formatvorgaben zur Manuskriptgestaltung. Dabei hatten wir mit Verwunderung deutliche wie konstante, teilweise abenteuerliche Abweichungen von den Vorgaben zur Manuskriptgestaltung feststellen müssen. Nahezu alle eingereichten Abstracts wiesen in der ersten Runde mehr oder weniger umfangreich formale Mängel auf. Diese Mängel führen wir im Nachhinein u.a. auf Probleme der Kenntnis, der Akzeptanz und des praktischen Beherrschens solcher Richtlinien zurück, zumal wenn diese relativ neu und teilweise auch (noch) unüblich sind. Praktisch war es zumeist ökonomischer und mit weniger Ärger für alle Beteiligten verbunden, aufgetretene Mängel (sofern sie nicht offensichtlicher Achtlosigkeit geschuldet schienen) beim Organisationskomitee bzw. bei der Herausgeber-schaft selbst zu korrigieren und kein aufwendiges Rückkoppelungsverfahren zu etablieren, das seinerseits der Kontrolle bedurfte. Im übrigen steht zu erwarten, dass sich dieses Problem im zunehmenden praktischen Umgang mit den neuen Richtlinien mittelfristig entschärfen wird. Voraussetzung dafür ist allerdings die Kenntnisnahme und Akzeptanz solcher Vorgaben, die wir trotz vielfältiger Hinweise (noch) nicht durchgängig feststellen konnten.
- Am Rande sei noch erwähnt, dass die Abwicklung des Review-Verfahrens über das Netz (via Email) deutliche organisatorische Vorteile bietet. Beim Sportspiel-Symposium traten hier einige Schwierigkeiten auf. Neben technischen Problemen kamen noch Ortsabwesenheiten von Referentinnen und Referenten etc. hinzu. Damit waren Irritationen und Missverständnisse, aber auch zeitliche Verzögerungen und erhöhter organisatorischer Aufwand vorprogrammiert. Hier scheint ein ‚notification‘-Verfahren, d.h. eine (Rück-)Bestätigung des elektronischen Schriftverkehrs angezeigt, um die Kommunikation zu sichern und zu verbessern.
- Insofern dieses Review-Verfahren formal ein ‚generelles‘ Verfahren war, das eine ‚unterschiedslose‘ Anwendung fand, stellt sich zwangsläufig die Frage nach dem Umgang mit Besonderheiten bzw. dem ‚Einzelfall‘. Es zeigte sich, dass eine Struktur von Evaluationskriterien, die eher auf Arbeiten mit quantitativen Schwerpunkten bezogen ist, Beiträgen mit eher qualitativen Schwerpunktsetzungen oft nicht hinreichend gerecht werden kann. In Konsequenz wurde ein Vorschlag zu einer differenzierten Begutachtung gemäß spezifizierten Kriterien (vgl. 4) entwickelt.
- Nicht hinreichend lösbar scheint uns derzeit eine zeitliche Problematik: wenn zum Zeitpunkt der Abfassung eines Abstracts Untersuchungen noch andauern, lassen sich im Abstract „Ergebnisse“ weder darstellen noch diskutieren und entsprechend auch begutachten. Dies unterscheidet – neben einigen anderen Kriterien – Tagungsbeiträge von solchen in

wissenschaftlichen Zeitschriften. Ob eine derartige Unbestimmtheit Qualitätsmängel impliziert, können und wollen wir an dieser Stelle nicht abschließend beurteilen. Auf entsprechende Beiträge zu verzichten, hieße jedoch auch, auf Aktualität zu verzichten.

Zusammenfassung

Grundlagen einer wissenschaftlichen Qualitätssicherung von Tagungsbeiträgen lassen sich mit einem Peer-Review-Verfahren dann erreichen, wenn folgende Bedingungen erfüllt werden (können): Das Verfahren muss in jeder Hinsicht seriös, d.h. vertraulich wie konstruktiv, aber auch transparent und gerecht für alle Beteiligten sein. Es bedarf in zeitlicher wie in technisch-organisatorischer Hinsicht einer angemessenen Form, die auch Abweichungen, Brüche und Widersprüche berücksichtigen kann. Ein solches Verfahren ließe sich jedoch auch als Gängelung und Normierung wissenschaftlichen Arbeitens, als Infragestellen und als Kritizismus persönlicher wissenschaftlicher Leistung, insofern als Belastung, Zumutung oder gar als Beleidigung verstehen. Dann wird das Verfahren scheitern. Das Verfahren und die Rückmeldungen lassen sich aber auch als – wie auch immer – konstruktiv, hilfreich, anregend usw. begreifen. Dann hat das Verfahren einen guten Teil seines Zwecks erfüllt.

4 Konsequenzen: Ein Vorschlag und ausgewählte Perspektiven

Das dargestellte Review-Verfahren war (auch) Diskussionsgegenstand auf der Sitzung der dvs-Kommission Sportspiele. Die Kommission einigte sich auf eine Reflexion und Optimierung des (prinzipiellen) Verfahrens und bat das (Bremer) Organisationskomitee, einen entsprechenden Vorschlag zu entwickeln und zur Diskussion zu stellen.

Wir haben darauf hin

- a) die von uns verwendeten Evaluationskriterien einer Itemanalyse unterzogen und
- b) Möglichkeiten eines Kriterienkatalogs für (Tagungs-) Beiträge mit qualitativen und quantitativen Schwerpunktsetzungen untersucht.

Die Ergebnisse werden im folgenden vor- und zur Diskussion gestellt.

Ad a) Im ersten Schritt wurde Signiererobjektivität, d.h. die Übereinstimmung der beiden Gutachter/innenurteile für die eingereichten Beiträge bestimmt. Hierfür wurde der gewichtete Kappa-Koeffizient für die einzelnen Items, z.B. Originalität, Darstellung der Ergebnisse, Klarheit berechnet. Die Übereinstimmung der Signiererurteile kann für alle Items als unzureichend bezeichnet werden (Cohen's Kappa: $\kappa \leq 0.14$, $p \geq .32$). Aufgrund der unzureichenden Objektivität wurde eine Reliabilitätsprüfung nicht durchgeführt.

Ad b) Grundlage für die Entwicklung eines Vorschlags war die Ermittlung von Kriterien, die (auch) international Anwendung finden. Wir haben uns dabei auf Begutachungskriterien von Peer-Reviews folgender Zeitschriften gestützt: ‚*Journal of Teaching in Physical Education*‘ (USA), ‚*Spectrum der Sportwissenschaften*‘ (Österreich), ‚*Zeitschrift für Erziehungswissenschaft*‘ (Deutschland) und ‚*psychologie und sport*‘ (Deutschland). Dabei sind wir uns bewusst, dass sich Beiträge in wissenschaftlichen

Zeitschriften durchaus von Tagungsbeiträgen bzw. deren Abstracts unterscheiden (können). Unbestimmt bleibt dabei allerdings, ob und an welchen Kriterien wissenschaftlicher Qualität sich eine solche Differenz festmachen lässt, zumal es in jeder Hinsicht auch Akzentuierungen und Unterschiede gibt, geben kann und vielleicht auch geben sollte. Letztlich ist dies eine Frage der Konvention.

Bei den gutacherlichen Einschätzungen von Beiträgen zeichnen sich gemäß den Kriterien der Zeitschriften zwei Verfahrensweisen ab:

- eine (vorgegebene) semantische Differenzierung der jeweiligen Qualität, ggf. versehen mit der Möglichkeit weiterer, abweichender und/oder differenzierender Etikettierung sowie mit Begründungshinweisen sowie
- eine skalierte Einschätzung analog einem ‚Bewertungssystem‘, wobei die ‚Noten‘ mit Umfängen einer Überarbeitung korrespondieren.

Durchgängig erfolgt eine Gesamtbewertung qualitativer Art. Während die erste Verfahrens-Variante Vorteile sprachlicher Differenzierungsmöglichkeiten bietet, erleichtert die zweite Variante einen Vergleich von Beurteilungen gemäß der jeweiligen Quantifizierungen (wenngleich diese aus unterschiedlichen Gründen resultieren können). In beiden Varianten ist eine wiederholte Begutachtung für solche Beiträge vorgesehen, in denen größere Veränderungen notwendig sind, die als Auflage für eine Annahme geltend gemacht werden.

Atmosphärisch wird von (viel) Zustimmung zu ‚Peer-Reviews‘ bei ‚positiv‘ Begutachteten, von Enttäuschung bei ‚negativen‘ Gutachten von den Herausgebern/innen berichtet. Tendenziell scheint sich aber Akzeptanz im Sinne von konstruktiver Aneignung der Gutachten durchzusetzen.

Für das Beurteilungsverfahren von Tagungsbeiträgen mit qualitativen und quantitativen Schwerpunktsetzungen schlagen wir vor, grundsätzlich eine skalierte Einschätzung vorzunehmen. Damit soll die Vergleichbarkeit von Gutachten erleichtert und das Verfahren bei einer möglichen Vielzahl von Tagungsbeiträgen ökonomisiert werden. Eine Gesamteinschätzung a) im Sinne einer Annahme oder Ablehnung des Beitrags, und b) nach dem Umfang von Überarbeitungsaufgaben ist unverzichtbar. Zusätzlich sollte die Möglichkeit für weitere qualitative Einschätzungen, semantische Differenzierungen und zusätzliche Kommentierungen eingerichtet werden. Insbesondere die Kommentierungen bilden einen Schwerpunkt in der konstruktiven Evaluation, da einstimmig diese Hinweise als besonders hilfreich anzusehen sind.

Vor diesem Hintergrund schlagen wir als Kriterienkatalog für eine Annahme sportwissenschaftliche Tagungsbeiträge (bzw. deren Abstracts) mit qualitativen und quantitativen Schwerpunktsetzungen einen Evaluationsbogen vor (siehe S. 33).

Wir möchten diesen Vorschlag in der sportwissenschaftlichen ‚Scientific Community‘ zur Diskussion stellen. Dazu haben wir den Evaluationsbogen unter dem URL www.sport.uni-bremen.de/Sportspiele/Evaluation eingepflegt und eine separate Emailadresse eingerichtet. Wir möchten alle Kolleginnen und Kollegen auffordern, möglichst zahlreiche, vollständige und anregende Rückmeldungen zu schicken. Diese können/sollen auch über den Beurteilungsbogen hinaus gehen. Über die Ergebnisse dieser Diskussion werden wir im Frühjahr 2003 berichten.

Was erwarten wir von unserer Initiative?

Wir wollen an der grundsätzlichen Intention wissenschaftlicher Qualitätsentwicklung und -sicherung festhalten. Dazu benötigen wir ein ebenso akzeptiertes wie (bzw. weil) methodisch hinreichendes und praktikables Evaluationsdesign für Tagungsbeiträge. Als ein konsentiertes Instrument ließe es sich für weitere Tagungen innerhalb der dvs verwenden, ob nun nach Bedarf modifiziert oder auch weiter entwickelt. Eine weitere, sozusagen ‚alltägliche‘ Anwendungsmöglichkeit bietet sich bei der Beurteilung wissenschaftlicher Hausarbeiten, von Diplom- und Magister-Arbeiten bis hin zu Dissertationen. Gerade hier besteht ja nicht nur große Uneinheitlichkeit, sondern auch ein erhebliches Defizit an Transparenz, vor allem für die Qualifikant/innen.

Man mag Bestrebungen wie diese aber auch als überflüssig, als methodisch problematisch, als gängelnd etc. verstehen, – im Sinne einer Normierung und Standardisierung, die Kreativität einschränkt, Freiheit beschneidet, Abweichungen stigmatisiert. So schützt weder das beste Abstract vor einer mangelhaften sprachlichen bzw. medialen Präsentation eines Tagungsbeitrags, noch lässt sich Eloquenz, die einen Tagungsvortrag zur Sternstunde macht, in ein Abstract pressen. Gleichwohl sehen wir in unserem Vorschlag auch ein Potenzial: Indem Begutachtungen wissenschaftlicher Arbeiten kriterienbezogen und inhaltlich transparent gemacht werden, kann (sport-)wissenschaftlicher Nachwuchs daran lernen. Und wer glaubt, begutachtete Beiträge würden an wissenschaftlicher Kunst verlieren, sollte Locke (1990) lesen.

Wir freuen uns auf Ihre/Eure Rückmeldung!

Literatur

Locke, L. F. (1990). Why Motor Learning is Ignored: A Case of Ducks, Naughty Theories, an Unrequired Love. *Quest*, 42, 134-142.

Dr. Lutz Müller / Dr. Dirk Büsch
Universität Bremen
FB 9 – Studiengang Sportwissenschaft
Postfach 33 04 40
28334 Bremen
eMail: lmuller@uni-bremen.de
dbuesch@uni-bremen.de



sport goes media – Zwischen Tradition und Vision

16. Sportwissenschaftlicher Hochschultag der dvs

21.-23. September 2003 · Münster

Informationen unter www.dvs2003.de



**dvs-Symposium „Sportwissenschaft aktuell“
TT.MM.-TT.MM.JJJJ in Schlaraffenland**

Abstract-Evaluation

Autoren/innen _____

Titel _____

Gutachter/innen A / B

Bewertungsschema (Bitte ankreuzen): ausgezeichnet gut ausreichend ungenügend irrelevant¹

I.

1. Originalität des Themas

2. Sprachliche Qualität

3. Formale Qualität

4. Struktur des Beitrages

5. Theoretischer Bezugsrahmen

II.

6. Methode

Vorgehen ist verständlich

Methode ist der Fragestellung angepasst

7. Ergebnisse

Auswertung

Stichhaltigkeit in Bezug auf die Fragestellung

8. Diskussion

Angemessene Schlussfolgerungen

III.

9. Ich empfehle die

- (a) uneingeschränkte Annahme des Abstracts
- (b) Annahme mit geringfügigen Überarbeitungen
- (c) Annahme mit gravierenden Überarbeitungen
- (d) Ablehnung des Abstracts

10. Kommentare und Empfehlungen für die Überarbeitung des Abstracts

¹ ausgezeichnet: außerordentliche Qualität, keine Verbesserung notwendig
gut: gute Qualität, geringfügige Überarbeitungen empfohlen
ausreichend: geringe Qualität, vor Annahme gravierende Überarbeitung empfohlen
unzulänglich: mangelnde Qualität, Annahme nicht möglich
irrelevant: für diesen Beitrag irrelevant