

Die Beurteilung wissenschaftlicher Leistungen durch vergleichende Begutachtung von Udo Rempe, Kiel

(Hierzu Folie 2, den Foliennummern des Vortrags entsprechen die Seitennummern der pdf-Datei)

Im Wissenschaftsbereich spielen Beurteilungen insbesondere eine Rolle bei:

1. der Auswahl für Juniorprofessuren,
2. Zwischenevaluationen,
3. der Leistungsfeststellung und der Festsetzung von Leistungsentgelten,
4. der Programmakkreditierung und
5. der Prozessakkreditierung.

Die Leistungsbeurteilung hat sorgfältig im Rahmen des dienstlichen Ermessens zu erfolgen, da Artikel 33 Abs. 2 Grundgesetz vorschreibt:

„Jeder Deutscher hat nach Eignung, Befähigung und fachlicher Leistung gleichen Zugang zu jedem öffentlichen Amte“.

(Hierzu Folie 3)

Man benötigt Maße zur Beurteilung der Ergebnisse von Leistungsbeurteilungen.

Dann kann man auch die Güte verschiedener Verfahren erkennen.

Die Varianz von tatsächlichen Beurteilungsergebnissen kann in Prozent einer optimalen Varianz ausgedrückt werden.

(Hierzu Folie 4)

Ein Beurteiler kann alleine eine andere Person beurteilen.

Das ist die I-Beurteilung.

Ob dabei Leistungen oder nur Körperhöhen beurteilt werden, ist aus methodischer Sicht irrelevant.

Selbermachen fördert das Verstehen:

Ordnen Sie bitte spontan die Körperhöhe ihres linken Nachbarn einer von fünf Stufen zu:

- sehr klein,
- klein,
- normal,
- groß,
- sehr groß.

Schreiben Sie sich bitte das Ergebnis auf!

(Hierzu Folie 5)

Folgende Beurteilungsverfahren arbeiten vergleichend:

Das V-Verfahren vergleicht zwei Personen,

das Y-Verfahren vergleicht drei Personen,

das X-Verfahren vergleicht vier Personen.

Selbermachen fördert das Verstehen:

Beurteilen Sie die Körpergröße der beiden links von Ihnen sitzenden Nachbarn.

Geben sie der kleineren Person 0 Punkte!
 Vergeben Sie einen Punkt an den größeren Nachbarn!

(Hierzu Folie 6)

Zwei Beurteilungsexperimente zur Beurteilung der Körperhöhe lieferten folgende Gütemaße:

Bei der I-Beurteilung 18 %,

bei der vergleichenden X-Beurteilung 74 %.

Die Ergebnisse mit der X-Beurteilung sind nicht fehlerfrei, aber deutlich besser als bei der I-Beurteilung.

Tatsächliche Leistungsbeurteilungen liefern teils noch schlechtere Ergebnisse als das Experiment.

(Hierzu Folie 7)

Im einfachsten Fall ordnet ein Gutachter die Leistungen einer Person in eine der Stufen 1 bis 5 ein.

Die Beurteilung ist I-förmig:

Petrus



Adam

Bewertet wird mit Punkten, Zahlenwerten oder Worten.

Punkte	Leistung	Körperhöhe
0	sehr schwach	sehr klein
1	schwach	klein
2	normal	normal
3	gut	groß
4	sehr gut	sehr groß

(Hierzu Folie 8)

Bei der V-Beurteilung beurteilt ein Beurteiler zwei Personen.

Dem Beurteiler geben wir den Platzhalternamen **Petrus**.

Den zwei zu beurteilenden Personen die Platzhalternamen

Adam und **Eva**.

Petrus hat wie Sie für Ihre beiden Nachbarn 0 oder 1 Punkt zu vergeben, je nachdem, ob die Leistung geringer oder besser ist.

Die Rollen von Petrus, Adam und Eva könnten

Herr Pätzold, Herr Schulze und Herr Rempe spielen.

Sie könnten die Güte der heutigen Vorträge vergleichen.

(Hierzu Folie 9)

Wir lassen in Gedanken Herrn Pätzold, Herrn Schulze und Herrn Rempe an einem runden Tisch Platz nehmen.

Die drei bilden einen Beurteilungskreis.

(Hierzu Folie 10)

Auf dem Tisch liegt ein gelbes Petrus-Set.

Dort sitzt der Beurteiler.
Dort sitzt Herr Pätzold.
Es folgt im Uhrzeigersinn ein blaues Adam-Set.
Dort sitzt die erste zu beurteilende Person.
Dort sitzt Herr Schulze.
Links von ihm folgt ein rotes Eva- Set.
Dort sitzt die zweite zu beurteilende Person.
Dort sitzt Herr Rempe.
Jeder im Beurteilungskreis hat eine blaue und eine rote Beurteilungskarte.
Herr Pätzold schreibt auf die blaue Karte den Namen von Herrn Schulze.
Fand er Herrn Schulzes Vortrag besser als Herrn Rempes, schreibt er eine 1 auf die blaue Karte, sonst eine 0.

(Noch Folie 10)

Auf die rote Karte schreibt er den Namen von Herrn Rempe.
Ferner notiert er dort die verbliebene Bewertungszahl.
Die Bewertungszettel werden in eine Urne gelegt.

In diesem Kreis von zu Beurteilenden werden dann im Uhrzeigersinn die Sets um einen Platz weiter geschoben und dann die Rollen vertauscht wahrgenommen.
Nach einem erneuten Verschieben der Sets hat jeder einmal die „Petrus“-Rolle des Beurteilers wahrgenommen.

(Noch Folie 10)

Wir veranschaulichen das anhand der Körperhöhen von

1. Herrn Soll,
2. Frau Mehl und
3. Herrn Ober.

Wir gruppieren die drei Personen um den imaginären Beurteilungstisch.
Die um 100 Zentimeter verminderten Körperhöhen stehen jeweils bei ihren Namen.
Herr Soll vergleicht also die Körperhöhe von Frau Mehl mit jener von Herrn Ober.
Der kleineren Frau Mehl gibt er 0 Punkte auf der blauen Karte.
Dem größeren Herrn Ober gibt er 1 Punkt auf der roten Karte.

(Hierzu Folie 11)

Die Sets werden um einen Platz weitergerückt.
Frau Mehl vergleicht Herrn Ober und Herrn Soll.
Der größere Herr Ober erhält 1 Punkt auf der blauen Karte.
Der kleinere Herr Soll 0 Punkte auf der roten Karte.

(Hierzu Folie 12)

Die Sets werden nochmals verschoben.
Herr Ober spielt den Petrus.
Er vergleicht Herrn Soll und Frau Mehl.
Der größere Herr Soll bekommt 1 Punkt auf der blauen Karte.
Die kleinere Frau Mehl erhält 0 Punkte auf der roten Karte.

Damit hat jede Person einmal beurteilt.

Aber jede Person ist auch zweimal beurteilt worden.
Alle Personen haben einen blauen Adam- und einen roten Eva-Punkt.
Die beiden Punkte werden zu einer Punktesumme zusammengezählt.
Jede der möglichen Punktesummen von 0 bis 2 tritt einmal auf und gibt genau an,
wie viele Personen kleiner sind.

(Noch zu Folie 12)

Um die Redezeit einzuhalten wird auf Beiträge zur studentischen Beurteilung mit dem V-Verfahren und zur Anwendung des Y-Verfahrens bei Akkreditierungen verzichtet.
Sie sind getrennt bei anderer Gelegenheit zu behandeln.

(Hierzu Folie 13)

Beim X-Verfahren sind Vierergruppen vergleichend zu beurteilen. Den vier zu Beurteilenden geben wir die Platzhalternamen

Adam und **Eva**,
Maria und **Josef**.

Diejenigen, die Petrus beurteilen, bekommen die Platzhalternamen
Kain und **Abel**,
David und **Goliath**.

(Hierzu Folie 14)

Wir setzen einen Kreis von 13 zu Beurteilenden um einen imaginären Tisch - linkes Bild.

Auf diesem liegen

das **gelbe Petrus**-Set für den Beurteiler,
das **blaue Adam**- und das **rote Eva**-Set für zwei zu Beurteilende und
das **schwarze Maria**- und das **grüne Josef**-Set für zwei weitere zu Beurteilende.

(Hierzu Folie 15)

Wie vorzugehen ist, zeige ich anhand der Körperhöhe von 13 Personen.
Zuerst vergleicht Herr Soll als **Petrus** vier Personen:

1. Frau Gerd,
2. Herrn Meier,
3. Frau Holz und
4. Frau Beer.

Er hat

1. eine **blaue Adam**-Karte,
2. eine **rote Eva**-Karte,
3. eine **schwarze Maria**-Karte und
4. eine **grüne Josef**-Karte.

Diese ordnet er den Personen entsprechend der „Setfarbe“ zu.

Er notiert auf den Karten mit steigender Körperhöhe:

0, 1, 2 und 3 Punkte.

(Hierzu Folie 16)

Nach der Körperhöhe ergibt sich die Reihenfolge

1. Frau Beer,
2. Frau Holz,
3. Frau Gerd und
4. Herr Meier

In dieser Reihenfolge vergibt Herr Soll die Punkte 0, 1, 2 und 3.

Die Sets werden jeweils um einen Platz weiter geschoben.

Nachdem alle 13 Personen geurteilt haben, ist jede der Personen viermal beurteilt.

(Hierzu Folie 17)

So wie man beim V-Verfahren für jede Person zwei Beurteilungen erhielt, erhält man beim X-Verfahren für jede Person vier Beurteilungen.

1. blaue Adampunkte,
2. rote Eva-Punkte,
3. schwarze Maria-Punkte und
4. grüne Josef-Punkte.

Die vier Punkte liefern eine weiße Punktesumme.

Deren Werte liegen zwischen 0 und 12.

Sie geben an, wieviele von 12 Personen kleiner sind.

Bei Leistungsbeurteilungen zeigen sie, wie viele Personen leistungsschwächer sind.

(Hierzu Folie 18)

Die zweite blau markierte Zahlenzeile zeigt die anhand der Messwerte festgestellten Punktesummen.

Die vorletzte rote Zahlenzeile enthält die interessanteren Punktesummen, die nur durch reines Beurteilen ohne Kenntnis der Messwerte entstanden.

(Hierzu Folie 19)

Das rechte Punktediagramm zeigt auf der x-Achse die durch Beurteilen gefundenen Punktesummen und auf der y-Achse die Logarithmen der Körperhöhen.

Das Bestimmtheitsmaß erreicht den Wert 0,74.

Das linke Punktediagramm zeigt auf der x-Achse die mit der I-Beurteilung gefundenen Einstufungen und auf der y-Achse die logarithmierten Körperhöhen.

Das Bestimmtheitsmaß beträgt nur 0,02.

(Hierzu Folie 20)

Der Rangkorrelationskoeffizient beträgt bei der I-Beurteilung nur 0,08, bei der X-Beurteilung 0,67.

(Hierzu Folie 21)

Aus der Varianz der Punktesummen kann man auch ohne Kenntnis von Messwerten auf die Ungenauigkeit von Beurteilungen schließen.

(Hierzu Folie 22)

In einem Fünftel bis zu einem Viertel der Fälle ordnen zwei Gutachter die gleichen zwei Personen unterschiedlich nach der Körpergröße. Etwa in einem **Achtel** der Fälle hält ein Gutachter die kleinere von zwei Personen für die größere. Auch bei Leistungsbeurteilungen kann man die Ungenauigkeit ermitteln.

(Hierzu Folie 23)

Hier wird gezeigt, wie eine X-Beurteilung hochschulübergreifend gestaltet werden kann.

Zunächst werden die neun Universitäten nach dem Los im Beurteilungskreis angeordnet.

Innerhalb des Bereiches jeder Universität wird die Reihenfolge der 3 Personen ebenfalls durch das Los festgelegt.

Durch die Anordnung der Sets wird in diesem Fall sichergestellt, dass der erste externe Bonner Gutachter Personen aus vier verschiedenen Hochschulen, nämlich aus

1. Köln,
 2. Münster,
 3. Aachen und
 4. Duisburg
- vergleicht.

(Noch zu Folie 23)

Zwar kommt es darauf an, dass nach Möglichkeit, wie dies auf der rechten Seite dargestellt ist, die Abstände zwischen den **Adam**-, **Eva**-, **Maria**- und **Josef**-Sets vergrößert werden. Aber nur bei Einhaltung der im Anhang aufgeführten Regeln findet man die dort dargestellte richtige Anordnung.

Man kann jedoch auch ausgehend von den auf Folie 14 vorgestellten Grundanordnungen für 13 (links) und 14 (rechts) Personen (Bei 14 Personen ist kein Vergleich mit 12 anderen Personen möglich!) ohne Verstoß gegen diese Regeln etwas suboptimale Anordnungen finden, wenn man jeweils zwischen **Kain** und **Petrus** und zwischen **Petrus** und **Adam** in gleicher Anzahl die erforderlichen unmarkierten Plätze einfügt.

(Noch Folie 23)

Auch bei allen anderen vergleichenden Gutachten werden in diesem günstigen Fall nur die Leistungen von Personen aus vier verschiedenen Hochschulen bewertet. Die mittleren Punktesummen der Personen einer Hochschule bewerten auch deren Fachbereich.

Sie liefern der für die Programmakkreditierung zuständigen Akkreditierungsagentur die nötigen Informationen darüber, ob die verlangte Qualität von Lehre und Forschung sowie des Qualitätsmanagement eines Fachbereichs in einem angemessenen Verhältnis zueinander stehen.

(Hierzu Folie 24)

Es wäre gut, wenn solche Beurteilungen für Hochschullehrerinnen und Hochschullehrer und der anderen Wissenschaftlerinnen und Wissenschaftler länderübergreifend erfolgen könnten.

Aber auch Pilotprojekte innerhalb eines großen Bundeslandes wie NRW wären bereits ein Fortschritt.

Sie, Herr Rudinger, könnten in der Psychologie,

Sie, Herr Hornbostel, in der Soziologie und

Sie, Herr Grüneberg, im Ministerium die Voraussetzungen dafür schaffen.

(Hierzu Folie 25)

Für eine länderübergreifende Einführung wäre eine Zusammenarbeit zwischen einer Akkreditierungsagentur und den Fachbereichs- oder Fakultätentagen wünschenswert. Zielsetzung müsste sein:

1. fachgebundene Qualifikationsrahmen unter Beachtung der Zielsetzungen des allgemeinen deutschen Qualifikationsrahmens,
2. die kombinierte Programmakkreditierung für die Studiengänge aller Fachbereiche und
3. die Organisation der Beurteilung individueller Leistungen.

(Hierzu Folie 26)

Zum Abschluss lade ich Sie zur X-Beurteilung ein.

Bitte nehmen Sie Platz!

Die Petrus-, Adam-, Eva-, Maria- und Josef-Sets liegen bereit.

Oder lassen Sie es.

Aber beklagen Sie sich dann nicht über mangelhafte Beurteilungen wissenschaftlicher Leistungen!

Anhang:

Bei der Anordnung der markierenden Sets sind nur bestimmte Abstände zwischen den durch Sets dargestellten Markierungen zulässig. Diese Abstände werden so benannt, dass sie von einer zuerst genannten Markierung im Uhrzeigersinn zur an zweiter Stelle genannten Markierung laufen.

Zwischen den vier zu Beurteilenden gibt es

1. einen kleinen Abstand k , der auf Folie 23 zwischen Adam und Eva liegt und dort den Wert $k=4$ hat,
2. einen mittleren Abstand m , der auf Folie 23 zwischen Maria und Josef liegt und dort $m=5$ beträgt, sowie
3. einen großen Abstand g , der auf Folie 23 zwischen Eva und Maria den Wert $g=7$ hat.

Diese Abstände müssen unabhängig davon, zwischen welchen beiden Markierungen sie liegen, verschieden sein.

Der große Abstand muss aber auch größer als die beiden anderen zusammen sein. Das ist im Schema auf Folie 23 nicht der Fall.

Zusätzlich darf der Abstand AJ von Adam über Eva und Maria bis Josef bestimmte Höchstwerte nicht überschreiten. Unkritisch sind Fälle, bei denen $(2 \bullet AJ + 1)$ kleiner als

die Gesamtzahl n der zu beurteilenden Personen ist. $2 \bullet AJ = n$ ist unzulässig. $2 \bullet AJ > n$ zulässig, wenn $2 \bullet g + 2 \bullet k + m < n$ und der kleine Abstand zwischen Eva und Maria liegt. Soll die Fremdbeurteilung gesichert werden, darf der Abstand PA von Petrus bis Adam nicht größer als die Anzahl p der zu Beurteilenden der gleichen Hochschule oder allgemeiner ausgedrückt der gleichen Gruppe sein.

Soll auch verhindert werden, dass ein Beurteiler seinen eigenen Beurteiler beurteilt, darf keine der Markierungen für Adam, Eva, Maria und Josef mit einer Markierung für Kain, Abel, David und Goliath zusammenfallen. Ferner ist der Abstand von Kain zu Petrus dann gleich dem Abstand von Petrus zu Adam und darf ebenfalls nicht den Wert p überschreiten.

Benutzt man für die Symbole \bullet , \bullet , \bullet , \bullet , \bullet und \circ für die Petrus-, Adam-, Eva-, Maria-, Josef und Nicht-Markierung, lässt sich ein besonders geeignetes Muster, das diese Vorschriften berücksichtigt, für die Set-Anordnung eines Beurteilungsrings in folgender Art platzsparend darstellen:

```

o o o o o  $\bullet$  o o o o o o o o
 $\bullet$ 
o o o o o  $\bullet$  o o o o  $\bullet$   $\bullet$ 

```

Die Vorzüge dieser Anordnung werden sichtbar, wenn man Petrus $\times * + \bullet$ von Kain, Abel, David und Goliath beurteilen lässt und die von ihnen mitbeurteilten Personen durch die vier unterschiedlichen Symbolgruppen (\times , \times , \times), ($*$, $*$, $*$), ($+$, $+$, $+$) und (\bullet , \bullet , \bullet) kennzeichnet.

Düsseldorf	Düsseldorf	Düsseldorf	Bielefeld	Bielefeld	Bielefeld	Bonn	Bonn	Bonn	Köln	Köln	Köln	Münster
\bullet	\circ	\bullet	\circ	$+$	\circ	$\times *$	\circ	$*$	\circ	$+$	\circ	$*$
\circ	\circ	$*$	\circ	$+$	\times	\circ	\circ	\bullet	\times	\circ	\times	\circ
Duisburg	Duisburg	Duisburg	Dortmund	Dortmund	Dortmund	Aachen	Aachen	Aachen	Wuppertal	Wuppertal	Wuppertal	Münster

Die erste Person aus Bonn wird mit 12 Personen aus allen 8 anderen Universitäten verglichen. Nur eine der 3 Bonner Personen ist auch Vergleichsperson. Bei der 2. Bonner Person würde überhaupt kein Vergleich mit einer Bonner Kollegin oder einem Bonner Kollegen erforderlich. Die Vergleichspersonen werden durch das veränderte „Set-Muster“ sehr gleichmäßig über den Beurteilungskreis verteilt.

Danksagung: Die Beurteilung der Körperhöhen von je vier Personen erfolgte durch Mitglieder des Landesvorstands und der Geschäftsstelle von dbb beamtenbund und tarifunion, Landesbund Schleswig-Holstein. Ohne deren Mitwirkung, für die ich mich bedanke, wäre diese Studie nicht möglich gewesen. Die Verantwortung für diesen Beitrag verbleibt voll bei mir. Die Namen der Beteiligten wurden entsprechend dem Wunsch eines Betroffenen in der schriftlichen Fassung durch Kunstnamen ersetzt.