

Aus der Serie "Doktorarbeiten mit methodischen Mängeln": Max und Murat

Beitrag von „kleiner gruener frosch“ vom 25. Juli 2018 19:00

In den letzten Wochen wurde wieder mal eine nette Studie aus einer Doktorarbeit veröffentlicht. Es ging um die unterschiedliche Beurteilung von "Max" und "Murat" in einem Diktat. Hier ein Link auf den Artikel auf <https://www.welt.de/vermisches/ar...te-als-Max.html>.

Schlussfolgerung der Studie aus einer Veröffentlichung der Doktorantin:

Zitat

Die Ergebnisse beider Studien deuten an, dass die angestrebte Leistungsgerechtigkeit im Hinblick der Bewertung von Schülerinnen und Schülern mit Migrationshintergrund noch nicht erreicht wurde. Die gute Nachricht ist allerdings, dass Lehrkräfte etwas dagegen tun können. So können beispielsweise im Vorhinein verschriftlichte Kriterien zur Zuordnung von Leistung zu Noten (z. B. Fehlertabellen) helfen, eine gerechtere Notenvergabepraxis zu erreichen.

Wie lief die Studie ab:

Zitat

Ein Teil der Studienteilnehmenden erhielt ein Diktat, welches angeblich von einem Schüler namens Max stammte, während der andere Teil ein Diktat von Murat erhielt. Die Fehlerzahl in beiden Diktaten war identisch und den Teilnehmenden gelang es in beiden Fällen gut, die Fehler zu identifizieren

Was lernen wir aus der Studie: "Verschiedene Lehramtsstudenten nutzen für die Bewertung von Diktaten eigene Fehlerquotienten und Bewertungsraster."

Was lernen wir aus der Studie nicht: "Verschiedene Kinder werden von ein und derselben Lehrperson unterschiedlich bewertet."

kopfschüttel

Beitrag von „Wollsocken80“ vom 25. Juli 2018 19:03

<https://www.lehrerforen.de/thread/47200-aus-der-serie-doktorarbeiten-mit-methodischen-m%C3%A4ngeln-max-und-murat/>

Nein, so ist es nicht. Ich kenne jemanden, der dort im Institut arbeitet, persönlich. Die Ergebnisse sind signifikant.

Beitrag von „Mikael“ vom 25. Juli 2018 19:11

Zitat

Nun wollen die Mannheimer Bildungswissenschaftler untersuchen, wie die Urteilsprozesse von Lehrern bei der Notenvergabe ablaufen und was die Gründe dafür sind, dass sie Schüler so unterschiedlich bewerten.

Nun ja, wäre das Ergebnis nicht "statistisch signifikant" wären die "Forscher" wohl jetzt arbeitslos. Aber so sieht man selbstverständlich "weiteren Forschungsbedarf"...

Gruß !

Beitrag von „Wollsocken80“ vom 25. Juli 2018 19:16

Was meinst Du mit "Forscher" in Anführungsstrichen? Es handelt sich hier um die Ergebnisse einer einzigen Doktorarbeit, die in einem grösseren Kontext zu sehen sind. So funktioniert Forschung in allen Fachbereichen. Die Ergebnisse sind signifikant, jedoch **nicht repräsentativ**, weil die Stichprobe zu klein war. Das geht aus der Studie aber auch eindeutig hervor. Der Spiegel erwähnt das sogar, während Die Welt das schon wieder unter den Tisch fallen lässt.

Beitrag von „Mikael“ vom 25. Juli 2018 19:17

Ohne die Studie im Detail zu kennen, sollten man beim Glauben an "statistische Signifikanz" sehr vorsichtig sein:

<https://www.spektrum.de/news/statistik...-fallen/1224727>

Gruß !

<https://www.lehrerforen.de/thread/47200-aus-der-serie-doktorarbeiten-mit-methodischen-m%C3%A4ngeln-max-und-murat/>

Beitrag von „Wollsocken80“ vom 25. Juli 2018 19:22

Und was willst Du damit sagen? Die Gruppe in Mannheim beschäftigt schon seit längerem mit dem Phänomen, das ist nicht die erste Studie die zu diesem Thema veröffentlicht wird. Reproduzierbarkeit ist hier kein Problem. Die Formulierung "deutet darauf hin" ist auch sehr vorsichtig und angemessen. Was die Medien daraus machen, ist ja eine andere Sache.

[Zitat von Mikael](#)

Ohne die Studie im Detail zu kennen

Wenn es Dich interessiert, kann ich Genaueres erfragen.

Bevor man übrigens über "methodische Mängel" schimpft, sollte man sich mal klar werden, was im Rahmen einer Dissertation möglich ist und was nicht. Auch der Tag eines Doktoranden hat nur 24 h und die Projektlaufzeiten sind in der Regel auf 3 Jahre begrenzt. Sämtliche Fehlversuche mit eingeschlossen natürlich.

Beitrag von „Mikael“ vom 25. Juli 2018 19:27

Ich will damit gar nichts sagen, eine Meinung über Sinn oder Unsinn der inflationär ansteigenden Zahl empirischer Studien im Bereich der [Pädagogik](#) kann sich ja jeder selbst bilden.

[Zitat von Wollsocken80](#)

Bevor man übrigens über "methodische Mängel" schimpft, ...

Das habe ich nicht getan, ich habe nur gesagt, dass man generell den Begriff "statistische Signifikanz" hinterfragen sollte, siehe den Spektrum-Artikel!

Gruß !

Beitrag von „Caro07“ vom 25. Juli 2018 21:31

Ich verstehe den Praxisbezug der Studie nicht. Wir haben schon seit Jahrzehnten eine Fehler - Notenzuordnung. Der gibt die Benotung vor. Gerade bei Diktaten kann man gar nicht ungerecht Fehler anstreichen.

Dass man Studenten die Fehler-Notenzuteilung frei wählen lässt, hat hier nichts mit dem Praxisbezug zu tun. Allerdings kann man an dieser fiktiven Situation schon Vorurteile herauslesen.

Eine ungerechte Benotung bei der Rechtschreibung wird in der Praxis nicht erfolgen, weil für die ganze Klasse ein Kriterienkatalog gilt. Die Schrift darf bei uns gar nicht in der Rechtschreibung bewertet werden, es sei denn, sie ist unleserlich, dann kann man das Wort nicht lesen und ist deswegen ein Fehler.

Beitrag von „kleiner gruener frosch“ vom 25. Juli 2018 21:46

Zitat

Allerdings kann man an dieser fiktiven Situation schon Vorurteile herauslesen.

Kann man natürlich. Man kann nur nicht mit Sicherheit sagen, dass Murat schlechter bewertet wird, weil er "Murat" heißt oder weil der Student kein Gefühl für die Bewertungsskala hat. Beides wäre denkbar.

Beitrag von „Jens_03“ vom 25. Juli 2018 21:47

Einfachste Erklärung:

Ein Teilnehmer hat 60 von 100 Punkten in einer [Klassenarbeit](#) erreicht.

IHK-Schlüssel = 3,9

Abikurse unserer Nachbarschule = 3,1

Uni-Notenschlüssel = 3,3

Beitrag von „Wollsocken80“ vom 25. Juli 2018 22:18

<https://www.lehrerforen.de/thread/47200-aus-der-serie-doktorarbeiten-mit-methodischen-m%C3%A4ngeln-max-und-murat/>

Ach Leute... Ihr glaubt ja wohl selber nicht, dass rein zufällig alle Studenten in der Murat-Gruppe genau den einen und die Studenten in der Max-Gruppe genau den anderen Bewertungsschlüssel verwendet haben. Bei einer randomisierten Zuteilung der Probanden auf die beiden Gruppen ist das gar nicht möglich.

Beitrag von „Anja82“ vom 25. Juli 2018 22:35

Hatte denn die Muratgruppe auch eine Maxarbeit und hat da bewusst den besseren Schlüssel benutzt? Wieso wählt man überhaupt unterschiedliche Schlüssel. Bei uns sind die vorgegeben, einer!

Beitrag von „kleiner gruener frosch“ vom 25. Juli 2018 22:35

[Wollsocken](#): Das glaubt ja auch keiner. Es waren bestimmt unterschiedliche Schlüssel/Maßstäbe, die jeder sich ausgedacht hat.

[@Anja](#), nein. Max-Gruppe hatte nur die Arbeit von Max.

kl. gr. frosch

Beitrag von „Jens_03“ vom 25. Juli 2018 22:38

Nein, aber MD/SD sind meiner Kenntnis nach bei ordinalskalierten Auswertungen nicht zulässig. Mal werden, laut Paper, die Probanden mit einem Aushang und durch persönlichen Kontakt rangeholt, dann wieder nur durch persönlichen Kontakt. Max. 50 Leute pro Kohorte sind nicht gerade viel und erklären die sehr geringen Unterschiede. Zumal die in der Welt falsch zitiert sind. Der Unterschied zwischen 1,87 und 2,03 sowie 3,64 und 4,15 ist nicht sehr groß (<https://www.frontiersin.org/files/Articles...-00481-t001.jpg>). Gehen wir von den Erkenntnissen der Sozialforschung aus, dann ist der Unterschied zwischen der 1,0 und der 4,4 geringer als der

zwischen der 4,4 und der 4,5 - das eine ist "bestanden", das andere ist "durchgefallen". Alles Punkte, die man berücksichtigen muss. Und das ist hierbei nicht geschehen.

Und eine gewisse Form von Notenschlüssel wird jeder im Kopf haben - bedingt durch die vorhandenen Vorerfahrungen.

Beitrag von „Wollsocken80“ vom 25. Juli 2018 22:46

Ja natürlich. Das ist ja auch die Quintessenz der Studie. Hier wird aber behauptet die Ergebnisse liessen sich allein darauf zurückführen und das ist gar nicht möglich. Die Probanden hatten sich eben nicht zuerst für den Bewertungsschlüssel entschieden und dann korrigiert, wie es sein sollte, sondern offensichtlich andersrum. Fehler haben sie ja auch alle gleich gezählt. Natürlich handelte es sich um eine fiktive Situation, die zeigen sollte, ob es Vorurteile gegenüber Migrantenkinder gibt. Wie im Welt-Artikel ja auch steht, kam die Gruppe zu ähnlichen Ergebnissen in einer Langzeitstudie mit 1500 Gymnasiasten in einer *realen* Situation. Ihr dichtet der Gruppe Dinge an, die sie selbst gar nicht für sich beansprucht.

Beitrag von „kleiner gruener frosch“ vom 25. Juli 2018 22:50

Zitat

Ihr dichtet der Gruppe Dinge an, die sie selbst gar nicht für sich beansprucht.

Die Aussage verstehe ich gerade nicht.

Beitrag von „Wollsocken80“ vom 25. Juli 2018 22:54

Es ist eine kleine Studie, die nach eigenem Statement der Gruppe nicht repräsentativ ist, weil die Stichprobe zu klein ist. Die müssen aber publizieren um weiter Geld zu bekommen, so läuft das immer und überall in der Forschung. Die Aussage ist "die Ergebnisse deuten darauf hin, dass..." und nicht "wir haben eindeutig gezeigt".

Beitrag von „Jens_03“ vom 25. Juli 2018 22:55

Problem bei der Auswertung sind unsere in Zahlen ausgedrückten Noten, die keinen "Wert" im Sinne dieser Publikation haben. Bsp. IHK Schlüssel:

"Sehr gut" = 100-92%

"Gut" = 91-82%

"Befriedigend" = 80-67%

usw.

Das ist keine Intervallskala. Und SD/MD für "3x sehr gut, 1x gut, 4x befriedigend" könnte man nicht bilden. Man muss diese Ergebnisse jedoch genauso auswerten und nicht über irgendwelche mathematischen Ansätze.

Daran hapert es. Und die Auswertung im Paper ist schlicht und ergreifend falsch. Das andere kenne ich nicht.

Beitrag von „Wollsocken80“ vom 25. Juli 2018 23:01

Um zu zeigen, dass überhaupt unterschiedlich bewertet wird, kann man die Daten natürlich so auswerten. Die Aussage, ob Murat damit nun durchgefallen oder versetzungsgefährdet wäre wird gar nicht getroffen. Gehen wir davon aus, dass es sich bei diesem fiktiven Diktat um eine reale Einzelnote handelt, kann es sehr wohl ausschlaggebend werden, ob Murat nun eine 3-4 oder eine 4 geschrieben hat.

Beitrag von „Jens_03“ vom 25. Juli 2018 23:17

Nein, eben nicht - zumal es den Unterschied zwischen 2 und 2 und 4 und 4 gibt. Und die angegebenen Werte deuten auf eine wesentlich höhere Spreitung in beiden Stichproben in Bezug auf die Noten hin. Dafür muss ich andere, naheliegender Methoden verwenden. Beispielsweise Mann-Whitney-U. Gebe ich dazu mal (fiktive) Zahlen mit vergleichbaren Mittelwerten, etc. ein, dann kommt man dort auf das Ergebnis "nicht signifikant". Das liest sich zu sehr, wie das im von Mikael geposteten Link: man sucht so lange, bis man einen signifikanten Unterschied findet. Würde die Studie strenger nach den Vorgaben der qualitativen Bildungs-/Sozialforschung ausgewertet werden, hätte ich damit weniger Probleme.

Beitrag von „Wollsocken80“ vom 25. Juli 2018 23:57

Der Unterschied beim besseren Diktat ist ja auch zu vernachlässigen, was im Übrigen gemäss Publikation der Erwartung der Autoren entspricht. Ich bin grad dabei mal die komplette Veröffentlichung zu lesen...

Beitrag von „kleiner gruener frosch“ vom 26. Juli 2018 00:18

Findet man die Publikation im Internet irgendwo?

Beitrag von „Sensei“ vom 26. Juli 2018 00:19

Was ich bei der Studie - zusätzlich zu den Ergebnissen - amüsant finde, ist auch die Tatsache, dass man, was Bewertungen im Schulalltag angeht, an der Uni ja relativ wenig lernt...das kommt doch oft erst im Referendariat oder im Job. Zu der realen Schulerfahrung der Partizipierenden:

Zitat von Studie

The participants in this study were 203 pre-service teachers (69.3% female) who were enrolled in a teacher training program at a university of education in Germany. They had a mean age of 23.39 ($SD = 3.42$) and a mean teaching experience of 2.12 months ($SD = 12.21$). All pre-service teachers were German and German native speakers. Within this sample, 86.8% of them had already successfully completed a school teaching internship as a mandatory part of their program.

The participants were recruited via notices posted on campus and through personal contacts. They received three Euros and chocolate for participating.

--> wir haben damals im Fachseminar auch eine Arbeit bekommen (EF, Deutsch, ohne EW) und in verschiedenen Gruppen wurde die Arbeit zwischen ausreichend und gut bewertet. Ist ja auch bei einer Textanalyse etc. viel einfacher als bei einem, wie hier in der Studie, vorliegendem Diktat von angeblichen Drittklässlern. Gerade ein Diktat eignet sich doch besonders wenig für eine solch variable Bewertung 🤔 Hier hätte man wirklich "sinnvoller" arbeiten können. Aber

vielleicht wären dann nicht diese in aktuellen Diskussion so wunderbar brauchbaren, polemisch verkürzten Ergebnisse herausgekommen. Und wer hinterfragt die Studie überhaupt? Letztendlich kursiert dann nämlich auch nur für die Jugend etc. auf Instagram [Sowas hier](#) - und gleich fühlen sich alle in ihren Reaktionen bestätigt und haben eigene Anekdoten für ihre Benachteiligung 🤪

Beitrag von „Wollsocken80“ vom 26. Juli 2018 00:42

Ähm... Gerade das Diktat eignete sich besonders gut um zu zeigen, was dann ja auch rauskam. Alles andere hätte ja nur zusätzliche Hintergrundvariablen implementiert. Was die Medien aus den Ergebnissen machen, dafür können die Leute nichts, die die Studie veröffentlicht haben. Auch die Auswahl der Probanden mit noch fast gar keiner Lehrerfahrung passte zur Fragestellung. Vielleicht einfach mal die ganze Studie lesen. Da alle auf dem Campus der Uni Mannheim rekrutiert wurden und eben so gut wie keine Lehrerfahrung hatten, erübrigt sich auch die Spekulation über die bewusst angewandten Notenschlüssel. Diese bewusste Entscheidung hat sicher nicht stattgefunden.

Beitrag von „Krabappel“ vom 26. Juli 2018 01:09

[Zitat von kleiner gruener frosch](#)

Wahnsinns Ergebnis. Hab ich noch nie gesagt, heute also zum ersten Mal: "und dafür zahlen wir Steuergelder?"

Beitrag von „Sensei“ vom 26. Juli 2018 01:14

Gerade ein Diktat mit einem - eigentlich - so eindeutigen Zusammenhang von Fehlerzahl und Note bietet sich hier ja nicht an - da gibt es keinen Spielraum, wenn man den EW dazugibt. Wenn man keinen Erwartungshorizont anbietet, kommt natürlich was Unterschiedliches heraus, wenn nicht einmal die Kriterien genannt werden außer "welche Note würdest du denn hier so gefühlstechnisch mal geben". Komplexe Hintergrundvariablen gehören nunmal bei unseren

pädagogischen Noten dazu 😊

Zitat von Studie

The participants were asked to rate the performance of the shown dictation by giving it a grade and counting the number of errors (dependent variable). The participants could enter the errors and the grade in an open field. ("How many mistakes did the dictation have?" and "What grade would you award the student for this dictation?"). They were asked to apply the German grading system (range from 0.75 to 6.00 with 0.75 indicating the best performance and 6 the worst performance) when grading and to count the mistakes in terms of the errors they found.

--> das hat meines Erachtens nichts mit schulischem Bewerten von Klassensätzen einer Arbeit zu tun. Deshalb hat diese Studie deutliches Potential für eine Optimierung.

Beitrag von „Mikael“ vom 26. Juli 2018 01:41

Interessant wäre ja auch einmal zu wissen, ob diese Forscher mit dem Beginn ihrer Forschung einfach so elfenbeinturmmäßig drauflosgeforscht haben, oder ob sie vorher beim Studiendesign ein paar Lehrkräfte mit Praxiserfahrung gefragt haben...

Gruß !

Beitrag von „Jens_03“ vom 26. Juli 2018 07:10

Zur Studie: <https://www.frontiersin.org/articles/10.33...00481/full#B28>

Die ursprüngliche Forschungsfrage zu verfolgen und dann zu sagen "Total egal, ob der Prüfling Murat oder Max heißt, die angehenden Lehrkräfte sind offenbar bei der Bewertung nicht ausländerfeindlich", ist offenbar nicht reißerisch genug. Zumal in einem echten Diktat Murat sich zu recht beschweren würde, wenn Max bei gleicher Punktzahl eine andere Note hätte.

Beitrag von „Valerianus“ vom 26. Juli 2018 08:30

<https://www.lehrerforen.de/thread/47200-aus-der-serie-doktorarbeiten-mit-methodischen-m%C3%A4ngeln-max-und-murat/>

Frontiers ist eine brauchbare psychologische peer-reviewed Zeitschrift und Oliver Dickhäuser (der Betreuer der Dissertation) habe ich bisher noch nie als Meister des Unsinns erlebt. Ich lese mir das später mal in Ruhe durch.

2 Anmerkungen:

[@Jens_03](#): Natürlich kann man einen Mittelwert aus Noten bilden, das wäre der Median.

[@Mikael](#): Natürlich wäre es für die universitäre [Pädagogik](#) deutlich besser, wenn weiterhin einfach irgendwelche Leute ihre Theorien publizierten ohne diese jemals auf ihren Wahrheitsgehalt zu prüfen. Auch dein Artikel tut so als wäre er etwas weltbewegend Neues, was noch nie einem Mathematiker oder Psychologen in den Sinn gekommen wäre. Das dort genannte Problem könnte man in 99% der Fälle durch Bonferroni-Korrektur beheben, der Grund warum das nicht passiert liegt nicht in fehlendem Wissen darum, sondern in der Angst dann weniger publiziert zu bekommen, dadurch weniger Forschungsgelder zu bekommen und dann weg vom Fenster zu sein. Also ein Fehler in der Finanzierung, nicht im statistischen Wissen. Brauchst du noch Artikel aus richtigen Fachzeitschriften oder glaubst du mir auch so, dass du da bei fast allen empirisch Forschenden offene Türen einrennst? 🤔

Beitrag von „Wollsocken80“ vom 26. Juli 2018 11:35

[Zitat von Sensei](#)

Gerade ein Diktat mit einem - eigentlich - so eindeutigen Zusammenhang von Fehlerzahl und Note bietet sich hier ja nicht an - da gibt es keinen Spielraum, wenn man den EW dazugibt.

Genau das war aber die Fragestellung und die hast Du offensichtlich nicht verstanden oder Du hast (wahrscheinlicher) die Publikation immer noch nicht vollständig durchgelesen. Nebenbei bemerkt ist es unhöflich, sich selbst "Sensei" zu nennen, die korrekte Bezeichnung wäre "Kyoushi".

[Zitat von Mikael](#)

Interessant wäre ja auch einmal zu wissen, ob diese Forscher mit dem Beginn ihrer Forschung einfach so elfenbeinturmmäßig drauflosgeforscht haben, oder ob sie vorher beim Studiendesign ein paar Lehrkräfte mit Praxiserfahrung gefragt haben...

Lies Dir doch die Publikation durch, dann weisst Du auch, welchen Aufwand die vor der eigentlichen Studie betrieben haben.

[Zitat von Krabappel](#)

Wahnsinns Ergebnis. Hab ich noch nie gesagt, heute also zum ersten Mal: "und dafür zahlen wir Steuergelder?"

Empirische Forschung im Bereich der [Pädagogik](#) verfolgt den Zweck, diesen Bereich aus der Welt der Binsenweisheiten und des hab-ich-immer-schon-so-gesagts herauszuholen. Wenn ich nun sage, meinem Gefühl nach könnte es den Garvorgang beschleunigen, die Kartoffeln bei 120 °C in einem Dampfdruckkochtopf zu kochen, muss ich das auch erstmal in einer sorgfältig geplanten Versuchsreihe zeigen um meine Hypothese als Allgemeingültigkeit verkaufen zu können. Zum Glück haben das in dem Fall bereits andere für mich erledigt, auf die ich mich berufen kann. Das ist gut, denn so muss ich das nicht jedes mal aufs Neue experimentell bestätigen, bevor ich es meinen Schülern erzähle. Es ist schon seltsam, dass diese Praxis im Bereich der Naturwissenschaften allgemein akzeptiert ist, im Bereich der [Pädagogik](#) aber ein ums andere mal bestenfalls belächelt und schlimmstenfalls verrissen wird.

Beitrag von „Anja82“ vom 26. Juli 2018 11:50

In der momentanen gesellschaftlichen Lage finde ich jedenfalls gefährlich mal oberflächlich rauszuposaunen, dass Lehrer Murat benachteiligen, wenn das gar nicht so klar ist.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 11:58

[Zitat von Anja82](#)

oberflächlich rauszuposaunen

Hat ja auch keiner. Hast *Du* denn die Publikation gelesen? Falls die Antwort "nein" ist, würde ich mich mal lieber zurückhalten, darüber oberflächlich zu urteilen.

Beitrag von „Anja82“ vom 26. Juli 2018 12:02

Glaubst du denn jeder der Welt-Leser macht das?

Ich lese eine Überschrift:

"Murat bekommt für dasselbe Diktat eine schlechtere Note als Max"

Ich lese den Eingangssatz:"in Diktat, gleich viele Fehler - doch die Lehrer vergeben dafür unterschiedliche Noten. Murat bekommt für identische Leistungen im Fach Deutsch eine schlechtere Beurteilung als Max. Das wiesen Forscher der Universität Mannheim in einer [experimentellen Studie](#) nach."

Und mehr liest doch der Welt-Leser nicht. Und das wird sicher nun auch weiter in andere Medien verteilt.

Und dann noch:

"Dass es überhaupt möglich ist, ein Diktat so unterschiedlich zu bewerten, liegt daran, dass es keine klaren Standards für die „Urteilsfindung“ der Lehrer bei einer Benotung gebe, erklärt Meike Bonefeld, Mitautorin und Leiterin der Auswertungen. Lehrer dürfen also beispielsweise im Diktat auch andere Faktoren als die Fehler heranziehen - beispielsweise das Schriftbild. Die Begründungen für die erstaunlichen Unterschiede in der Bewertung derselben Leistungen von „Max“ und „Murat“ seien aber nicht mehr nachzuvollziehen."

Das lesen die meisten schon nicht mehr. Ist aber eigentlich eine Falschaussage.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 12:08

[Zitat von Anja82](#)

Und mehr liest doch der Welt-Leser nicht.

Hier diskutieren wir aber unter Lehrern, also ausgebildeten Akademikern. Da erwarte ich schon, dass man sich mit der Sache differenzierter auseinandersetzt bevor man lospoltert. Wir Naturwissenschaftler müssen im Übrigen auch damit leben, dass die Medien alle Nase lang Studienergebnisse zu irgendwas aufblähen, das die veröffentlichenden Forscher für sich selbst gar nicht beanspruchen. So ungefähr jeder Ernährungsmythos lässt sich z. B. auf solche Aktionen zurückführen. Nachdem ich die Studie gelesen habe, denke ich, dass die Max-und-Murat-Geschichte erheblich seriöser dasteht als "Salz macht süchtig!". SPON stellt die Sache

übrigens deutlich differenzierter da, als die Welt. Darauf habe ich aber gestern schon mal hingewiesen.

Beitrag von „Anja82“ vom 26. Juli 2018 12:11

Auch wenn wir HIER unter Lehrern diskutieren, darf ich es dennoch nicht gut finden, in der aktuellen Lage sowas so ungefiltert und dann noch mit fehlerhaften Informationen rauszuposaunen.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 12:13

[Zitat von Anja82](#)

fehlerhaften Informationen

Kannst Du denn überhaupt beurteilen, ob die Informationen fehlerhaft sind ohne die Publikation gelesen zu haben?

Beitrag von „Anja82“ vom 26. Juli 2018 12:22

Ich urteile, dass die im Weltartikel fehlerhaft sind. Denn gerade bei Diktaten gibt es klare Standards. Ist Studenten im 3. Semester vielleicht nicht bekannt.

Beitrag von „Jens_03“ vom 26. Juli 2018 12:25

[Zitat von Valerianus](#)

[@Jens_03](#): Natürlich kann man einen Mittelwert aus Noten bilden, das wäre der Median.

Ist mir klar. Aus dem Kontext der Verwendung der Abkürzung M im Paper, mit bspw. der Angabe, dass beim Alter M=22,28 Jahre ist, bezweifle ich, dass der/die Teilnehmende Nr.XYZ zum Zeitpunkt der Untersuchung 22 Jahren und 102,27 Tage alt war. Naheliegender dürfte die Erfassung des Alters mit 22 Jahren, etc. sein, woraufhin M dann vermutlich das arithmetische Mittel darstellt.

Da die Autoren nicht deutlich machen, dass sie eventuell doch in einer Tabelle dann mit M den Median meinen, ebenso nicht angeben, ob die Teilnehmenden die Noten auf zwei Nachkommastellen angeben sollten (sonst kommen wir nicht auf den Median=1,87, etc.), gehe ich davon aus, dass wir eben keinen Median haben, sondern einen wie auch immer berechneten anderen Mittelwert.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 12:37

[Zitat von Anja82](#)

Ich urteile, dass die im Weltartikel fehlerhaft sind. Denn gerade bei Diktaten gibt es klare Standards. Ist Studenten im 3. Semester vielleicht nicht bekannt.

Dann urteilst Du falsch. Zitat aus der Welt:

"Das Team des Lehrstuhls Pädagogische Psychologie um Oliver Dickhäuser wollte wissen, ob **angehende** Lehrer die Leistungen von Schülern mit ausländischen Wurzeln anders beurteilen als jene mit deutschem Hintergrund."

Das ist korrekt. Die Fragestellung zielte nämlich **nicht** darauf ab, ob das Diktat von einer **erfahrenen** Lehrperson in der Realität und nach einem **festgelegten Notenschlüssel** ebenso mangelhaft beurteilt worden wäre. Ziel der Studie war es, genau das zu zeigen: ein festgelegtes Beurteilungsraster sollte es bei **jeder** Leistungskontrolle geben um genau das zu vermeiden, worauf die Probanden eben reingefallen sind. Am besten liess sich dies eben anhand eines Diktats zeigen, bei dem einmal Fehler gezählt werden (da gibt es eben nur richtig oder falsch, unabhängig vom Namen des Kindes) und einmal ohne vorher festgelegtes Bewertungsraster eine Note vergeben wird. Die Studie ist ganz bewusst so aufgezogen worden, das wüsstest Du, hättest Du sie eben gelesen.

Es ehrt Dich, dass Du immer nach einem festgelegten Bewertungsraster korrigierst. Wir hatten in diesem Forum aber nicht nur einmal die Diskussion darüber, dass das gerade bei Aufsätzen & Co. längst nicht von allen Kollegen so gehandhabt wird.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 12:38

[Zitat von Jens 03](#)

dass wir eben keinen Median haben, sondern einen wie auch immer berechneten anderen Mittelwert.

Soll ich nachfragen?

Beitrag von „Anja82“ vom 26. Juli 2018 12:50

Aufsätze kann ich nicht sagen, da ich kein Deutsch mehr unterrichte. In meiner Schule werden im Jahrgang gleiche Arbeiten geschrieben. In Ma/ SU z.B. legen wir genau fest wofür es Punkte gibt, was ein Fehler ist und bei wieviel Prozent man welche Note kriegt. Sollte mal Unklarheit herrschen (weil ein Kind mit seinem Fehler sehr kreativ war) besprechen wir uns.

Beitrag von „Krabappel“ vom 26. Juli 2018 12:50

[Zitat von Wollsocken80](#)

...

"Das Team des Lehrstuhls Pädagogische Psychologie um Oliver Dickhäuser wollte wissen, ob **angehende** Lehrer die Leistungen von Schülern mit ausländischen Wurzeln anders beurteilen als jene mit deutschem Hintergrund."

Das ist korrekt. Die Fragestellung zielte nämlich **nicht** darauf ab, ob das Diktat von einer **erfahrenen** Lehrperson in der Realität und nach einem **festgelegten Notenschlüssel** ebenso mangelhaft beurteilt worden wäre. Ziel der Studie war es, genau das zu zeigen: ein festgelegtes Beurteilungsraster sollte es bei **jeder** Leistungskontrolle geben um genau das zu vermeiden, worauf die Probanden eben reingefallen sind. ...

Nenne mir bitte einen schlüssigen Grund, warum man Lehramtsstudenten nachweisen muss, dass sie noch nicht wissen, wie man Diktate ordentlich auswertet. Hätten sie die Diktate angehenden Metzgern vorgelegt, wäre es mir recht gewesen. Nur so wird suggeriert, Lehrer seien halt prinzipiell unreflektiert.

Wenn dann noch die "gute Nachricht" verkündet wird, dass Lehrer was gegen ungerechte Noten tun können, nämlich nicht aus dem Bauch heraus Noten zu verteilen, dann kann ich mich schon ärgern. Studien, die nur aus dem Selbstzweck geschmiedet werden, damit irgendwer Forschungsgelder bekommt sind doch ärgerlich. Unklar, was man daran gut finden kann.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 13:08

[Zitat von Krabappel](#)

Nenne mir bitte einen schlüssigen Grund, warum man Lehramtsstudenten nachweisen muss, dass sie noch nicht wissen, wie man Diktate ordentlich auswertet.

Darum ging es nicht in der Studie. Mehr ist dazu jetzt auch nicht mehr zu schreiben.

[Zitat von Anja82](#)

In Ma/ SU z.B. legen wir genau fest wofür es Punkte gibt, was ein Fehler ist und bei wieviel Prozent man welche Note kriegt.

Ja, fein. In Chemie benote ich auch zu 90 % nach festgelegten Bewertungsrastern. Es gibt aber einige Formate der Leistungskontrolle, bei denen das nur bedingt möglich ist. Vorträge z. B., die ich aus genau diesem Grund auch gar nicht gerne bewerte. Da soll ich dann auf meinem schönen Raster Punkte für so lustige Dinge wie "Ausdruck" und "Präsenz" geben. Hurra. Wer als Lehrer behauptet, Subjektivität würde bei der Notenfindung keine Rolle spielen, der belügt sich selbst.

Beitrag von „Jens_03“ vom 26. Juli 2018 13:18

[Zitat von Wollsocken80](#)

Soll ich nachfragen?

Gerne.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 13:22

Ich versuche es. Ich kenne wie gesagt jemanden aus dem Umfeld der Gruppe, nicht aber die Autorin selbst. Kann aber nen Moment dauern.

Beitrag von „Jens_03“ vom 26. Juli 2018 13:39

[Zitat von Wollsocken80](#)

Ich versuche es. Ich kenne wie gesagt jemanden aus dem Umfeld der Gruppe, nicht aber die Autorin selbst. Kann aber nen Moment dauern.

Ich bin gespannt.

Die Forschungsfrage an sich ist interessant. Und ich habe genug Kolleg(inn)en vor Augen, die (aus welchen Gründen auch immer) äusserst subjektiv bewerten. Ich kann auch die Forschenden verstehen. Da hat man eine schöne Forschungsfrage, 609 € ausgezahlt plus die Schokolade, vielversprechende Vorstudien durchgeführt - und dann kommt doch nicht das Gesuchte dabei raus. Damit die Arbeit nicht umsonst war, sucht man weiter. Dazu noch die Situation der Lehrstuhlinhaber mit den W(eniger)-Stellen, die oft nur dann einen Aufschlag bekommen, wenn die Zahl der Publikationen über dem Schnitt der Hochschule/Fakultät/was auch immer liegt.

Das die Zeitungen daraus eine tolle Schlagzeile machen, die den Inhalt der Studie entstellt, liegt nicht bei den Autoren - da können die nichts zu. Aber das Paper wirft bei mir eben doch ein paar Fragen auf.

Beitrag von „Krabappel“ vom 26. Juli 2018 13:43

[Zitat von Wollsocken80](#)

Darum ging es nicht in der Studie. Mehr ist dazu jetzt auch nicht mehr zu schreiben.

Das ist allen bewusst, die Frage ist aber, was mit den Ergebnissen passiert, die in der Öffentlichkeit breitgetreten werden. Und sinnlos bleibt sie allemal, denn Lehrer wissen nunmal, welche Kriterien zur Bewertung herangezogen werden müssen.

[Zitat von Wollsocken80](#)

... Wer als Lehrer behauptet, Subjektivität würde bei der Notenfindung keine Rolle spielen, der belügt sich selbst.

Auch das bezweifle ich als Allerletzte. Lustig, bei der Diskussion um die Unabhängigkeit und Objektivität von Bildungsempfehlungen wollten mir alle weismachen, wie neutral und unabhängig vom Sozialstatus der Kinder doch bewertet würde...

Ich finde die Frage, die hinter der Studie steht durchaus interessant. Noch viel interessanter wären aber neue Erkenntnisse und Lösungsansätze jenseits von Binsenweisheiten, die zur Grundausbildung gehören.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 14:00

[Zitat von Krabappel](#)

Auch das bezweifle ich als Allerletzte. Lustig, bei der Diskussion um die Unabhängigkeit und Objektivität von Bildungsempfehlungen wollten mir alle weismachen, wie neutral und unabhängig vom Sozialstatus der Kinder doch bewertet würde...

Nein ... so war das nicht. Ich habe die fragliche Diskussion gerade noch mal rausgesucht und zitiere exemplarisch mich selbst:

[Zitat von Wollsocken80](#)

Es bestreitet eigentlich keiner nachgewiesene Zusammenhänge. Das jeweilige Fazit ist nur ein anderes. Du sagst, das dreigliedrige Schulsystem ist für die Füchse, ich sage, es

ist gut. Du siehst ein systemisches Problem, ich sehe ein Problem bei den Eltern. Ich nenne auch keine Einzelfälle sondern sage, ich unterrichte an zwei Schulstufen und sehe keine "falsch einsortierten" SuS (und wenn dann eher zu hoch, als zu tief im Niveau). Das Phänomen müsste bei uns ja noch viel ausgeprägter sein, da die Übertrittsquote ans Gymnasium so viel kleiner ist. Meinst Du denn alle SuS bekommen automatisch bessere Noten nur weil noch mehr ans Gymnasium gehen? Das Problem mangelnder häuslicher Unterstützung bleibt doch, egal an welcher Schulstufe.

[Zitat von Krabappel](#)

Und sinnlos bleibt sie allemal, denn Lehrer wissen nunmal, welche Kriterien zur Bewertung herangezogen werden müssen.

Das wage ich zu bezweifeln. Selbst wenn theoretisch das alle wissen sollten, gibt es genug, die sich nicht dran halten. Und wie bereits erwähnt gibt es Formate der Leistungskontrolle, die für Subjektivität erheblich anfälliger sind, als andere.

Beitrag von „Anja82“ vom 26. Juli 2018 14:24

Bei meiner Tochter in der weiterführenden Schule zählen schriftliche Arbeiten nur noch 40 Prozent. Man sollte sich lieber über die übrigen 60 Gedanken machen, denn mündlich Bewertungen finde ich noch viel subjektiver.

Beitrag von „Yummi“ vom 26. Juli 2018 15:39

Multiple Choice und gut ist 😎

Ach geht auch nicht, denn der Murat versteht sinnbefreite Aussagen schlechter als Max.

Beitrag von „Wollsocken80“ vom 26. Juli 2018 16:44

Anfrage ist nach Mannheim geschickt und ist in Arbeit. Man äusserte sich zumindest schon mal wohlwollend ob des Interesses an der Auswertung und versprach sich später ausführlich zu melden. 😊

Beitrag von „Jens_03“ vom 26. Juli 2018 18:59

Zitat von Wollsocken80

Man äusserte sich zumindest schon mal wohlwollend ob des Interesses an der Auswertung und versprach sich später ausführlich zu melden. 😊

Na dann, sind wir doch mal gespannt.

Beitrag von „Sensei“ vom 26. Juli 2018 22:33

Der gute Fußumhüllende - schon lieb, wie du diese Studie verteidigst bzw. versuchst, ihre Vorteile im Forum verständlich zu machen. Löblich. Vielleicht bist du da auch von Sympathien und Empathie mit den Verfassern geprägt? Wir sind nie ganz sachlich und objektiv - wissen wir aber doch auch so.

Aber ich erkenne nach wie vor keinen Mehrwert - und die Teile, die ich gelesen habe, haben mich nicht angeregt, mehr zu lesen, da das Ergebnis eben so banal ist. Und das, was damit gemacht wird ärgert mich nunmal auch, wurde aber auch schon kommentiert. Und damit reiht es sich (sicher unfreiwillig) ein in die Reihe von Studien und Umfragen, die seltsame oder gut nutzbare Ergebnisse bringen (wie diese lustige Studie, die angeblich bestätigte, dass kleinere Lerngruppen keinen Vorteil brächten - als Testgruppe wurden, wenn ich mich recht entsinne, zwei Schüler aus einer großen Lerngruppe entzogen...und aha, keine Veränderung feststellbar).

Beitrag von „Meerschwein Nele“ vom 27. Juli 2018 13:05

Zitat von Sensei

wie diese lustige Studie, die angeblich bestätigte, dass kleinere Lerngruppen keinen Vorteil brächten - als Testgruppe wurden, wenn ich mich recht entsinne, zwei Schüler aus einer großen Lerngruppe entzogen...und aha, keine Veränderung feststellbar).

Das interessiert mich - hast du dafür einen Quellenverweis?

Beitrag von „BlackandGold“ vom 27. Juli 2018 13:50

[Zitat von Wollsocken80](#)

Anfrage ist nach Mannheim geschickt und ist in Arbeit. Man äusserte sich zumindest schon mal wohlwollend ob des Interesses an der Auswertung und versprach sich später ausführlich zu melden. 😊

Danke, das würde mich interessieren.

Insgesamt finde ich ja diese Tabelle aufschlussreich:

<https://www.frontiersin.org/files/Articles...-00481-t001.jpg>

Das Cohen's d für den Zusammenhang "Namen"- "Note" dabei ist ja durchaus als "schwach" (weil zwischen 0.2-0,5) zu bezeichnen.

Kombiniert mal mit dieser Abbildung:

<https://www.frontiersin.org/files/Articles...-00481-g001.jpg>

Da stellt sich heraus, dass Leute mit einer positiven Einstellung zu Migranten offenbar für die krassen Abweichungen relevant sind. Die Forscher bemerken ja auch:

"One reason could be that the teachers with a positive implicit attitude toward students with a migrant background who had to assess a student with a migrant background who performed poorly had high expectations because of their positive attitude toward students with a migrant background, and these expectations were disappointed by the dictation. Thus, they may have graded the dictation more severely because it did not fulfill their expectation. However, before putting forward alternative explanations, the unexpected direction of the interaction between performance level, migrant background, and implicit attitude should be replicated in future studies."

Darf man das mit dem "Harter Boden der Realität"-Effekt vergleichen? Ich finde persönlich, dass dieser Effekt der interessanteste Effekt der gesamten Studie ist... Denn der würde uns deutlich mehr Aufschluss über die Weiterentwicklung der Lehrerausbildung an Universitäten geben.

Beitrag von „benminor7“ vom 27. Juli 2018 13:55

Zitat von Mikael

Ohne die Studie im Detail zu kennen, sollten man beim Glauben an "statistische Signifikanz" sehr vorsichtig sein:

<https://www.spektrum.de/news/statistik...-fallen/1224727>

Gruß !

Ich will hier keine Klischees auffahren. Aber es kommt doch häufiger vor als man denkt, dass Hypothesen und Nullhypothesen nach den Untersuchungen den Ergebnissen angepasst werden - eben damit „Signifikanz“ entsteht...

Beitrag von „Jens_03“ vom 27. Juli 2018 13:55

Zitat von BlackandGold

Darf man das mit dem "Harter Boden der Realität"-Effekt vergleichen? Ich finde persönlich, dass dieser Effekt der interessanteste Effekt der gesamten Studie ist... Denn der würde uns deutlich mehr Aufschluss über die Weiterentwicklung der Lehrerausbildung an Universitäten geben.



Beitrag von „kleiner gruener frosch“ vom 27. Juli 2018 13:59

<https://www.lehrerforen.de/thread/47200-aus-der-serie-doktorarbeiten-mit-methodischen-m%C3%A4ngeln-max-und-murat/>

Hm, habe ich den Abschnitt richtig verstanden? Bei den "schlechteren Diktaten" geben die Personen mit einer gemessenen positiven Einstellung zu Murat ihm schlechtere Noten, weil sie (nach Interpretation der Forscher) enttäuscht von der Leistung von Murat sind und ihn das "spüren lassen wollen". Und die mit einer gemessenen negativen Einstellung zu Murat geben ihm tendenziell noch die bessere Note (alle so im Bereich der 4+).

Richtig? (Mein Englisch ist leider manchmal etwas wackelig - besonders bei Forschungsstudien.



kl. gr. frosch

Beitrag von „Mikael“ vom 27. Juli 2018 14:24

Fassen wir einmal zusammen:

- hochemotionales Thema auf das sich die MSM begeistert stürzen
- "signifikantes" Ergebnis
- viele offene Fragen und "weiterer Forschungsbedarf"

Ist doch optimal gelaufen für die Forscher...

Gruß !