

Gruppenleistungen beim Review von Multiple-Choice-Fragen - Ein Vergleich von face-to-face und virtuellen Gruppen mit und ohne Moderation

Zusammenfassung

Einleitung: Multiple-Choice-Fragen (MCF) werden in vielen Prüfungen der medizinischen Ausbildung verwendet und bedürfen aus diesem Grund einer sorgfältigen Qualitätssicherung, beispielsweise durch den Einsatz von Review-Komitees. Anhand der vorliegenden empirischen Studie soll erforscht werden, ob virtuell per E-Mail kommunizierende Review-Komitees vergleichbar sind mit face-to-face Review-Komitees hinsichtlich ihrer Leistung beim Review-Prozess und ob sich Moderation positiv auswirkt.

Methodik: 16 Kleingruppen von Psychologie-Studenten hatten die Aufgabe unter vier verschiedenen Versuchsbedingungen MCF zu bewerten und zu verbessern. Im zweiten Untersuchungsabschnitt wurden die veränderten Fragen zur Beurteilung der Güte einer neuen Stichprobe vorgelegt.

Ergebnisse: Für die Variablen „Art der Review-Komitees“ und „Moderation“ wurde kein signifikanter Einfluss gefunden. Jedoch unterschieden sich face-to-face und virtuelle Gruppen deutlich in den benötigten Bearbeitungszeiten. Die Versuchsbedingung „face-to-face ohne Moderation“ wurde von den Versuchsteilnehmern hinsichtlich Verantwortungsübernahme, Arbeitsweise, Wohlfühl, Motivation und Konzentration auf die Aufgabe generell am positivsten bewertet.

Diskussion: Face-to-face und virtuelle Gruppen sind beim Review von MCF gleich effektiv, unterscheiden sich aber hinsichtlich ihrer Effizienz. Der Einsatz elektronischen Reviews scheint somit möglich, ist aber wenig empfehlenswert wegen der langen Bearbeitungszeit und technischer Probleme.

Schlüsselwörter: Multiple-Choice-Fragen, face-to-face, virtuell, Moderation, Review-Komitee

Edda Kazubke¹

Katrin

Schüttpelz-Brauns²

1 Humboldt-Universität zu Berlin, Institut für Psychologie, Berlin, Deutschland

2 Charité - Universitätsmedizin Berlin, Dieter Scheffner Fachzentrum, Assessment-Bereich / Progress Test Medizin, Berlin, Deutschland

Einleitung

Die ärztliche Approbationsordnung verlangt für jedes Fach in der medizinischen Ausbildung eine Benotung. Zur Überprüfung der Studienleistung werden vielfach Multiple-Choice (MC) Aufgaben eingesetzt. Hierbei ist die Qualität der konstruierten Fragen von hoher Bedeutung. Zur Gewährleistung dieses Anspruchs ist es notwendig, die MC-Fragen vor ihrem Einsatz in Prüfungen zu evaluieren und gegebenenfalls zu verbessern. Die Qualitätssicherung erfolgt in der Regel durch den Einsatz von Review-Komitees (RKs), welche sich aus Experten verschiedener Fachbereiche konstituieren und bezüglich der Erkennung formaler Fehler in Fragen geschult werden. Diese spezielle Aufgabe wird in Gruppen bearbeitet, um das Wissen und die Erfahrung vieler Personen nutzbar zu machen. Um mögliche Störeinflüsse gruppenspezifischer Prozesse zu minimieren, werden die regelmäßigen Sitzungen moderiert. Moderation beeinflusst die Qualität von Kommunikation und Zusammenarbeit positiv [1]. Die Terminie-

rung der Mitglieder von RKs ist ein Problem, zu dessen Lösung die Einführung elektronischer Reviews beitragen könnte. Die Umsetzung virtueller Zusammenarbeit als RK ist beispielsweise per E-Mail denkbar. Von Vorteil ist hierbei die selbstbestimmte Wahl des Kommunikationszeitpunktes und der Bearbeitungszeit [2]. Die bisherige Forschung zum Vergleich der Arbeit von face-to-face und virtuellen Gruppen konnte keine Leistungsunterschiede zeigen. Virtuelle Gruppen benötigten allerdings deutlich mehr Zeit für die Bearbeitung derselben Aufgabe. Darüber hinaus waren die Mitglieder solcher Gruppen wesentlich unzufriedener mit der Zusammenarbeit. Ein positiver Einfluss von Moderation konnte auch für den Bereich virtueller Zusammenarbeit gezeigt werden [3]. Weitgehend unerforscht ist bisher die Vergleichbarkeit der Leistungen von face-to-face und virtuellen Gruppen bei der speziellen Aufgabenstellung des Review von MC-Aufgaben. Hieraus resultiert die Frage nach der Übertragbarkeit der Ergebnisse aus der sozialpsychologischen Forschung in den Bereich der medizinischen Ausbildung.

Zur Beurteilung der Leistung werden sowohl Effektivitäts-, als auch Effizienzkriterien herangezogen. Weiterhin soll untersucht werden, in wie weit Unterschiede zwischen face-to-face und virtuellen RK-Gruppen bezüglich der subjektiven Beurteilung von Verantwortungsübernahme, Zusammenarbeit, Wohlfühl, Motivation und Konzentration auf die Aufgabe existieren.

Methodik

Die Untersuchung erfolgte in zwei Abschnitten – der Durchführung von RKs zur Beurteilung und Verbesserung mangelhafter MC-Fragen und einem anschließenden Testlauf der veränderten Fragen zur Ermittlung der Güte.

Stichproben

Teilnehmer für die RKs waren 41 Studentinnen und 8 Studenten der Psychologie im Grundstudium der Humboldt-Universität zu Berlin mit einem Altersmittel von 23,17 Jahren ($SD=4,21$). Diese wurden durch Aushänge und direkte Ansprache in Vorlesungen rekrutiert und erhielten für ihre Teilnahme Versuchspersonen-Stunden vom Lehrstuhl für Sozialpsychologie gutgeschrieben. Sie arbeiteten in 16 Kleingruppen von zwei bis vier Personen. Die Anzahl der Gruppen wurde mittels G*Power festgelegt, um einen mittleren Effekt mit Hilfe der gewählten Teststatistiken belegen zu können. Die Hälfte der Gruppen war in ihrer geschlechtsbezogenen Zusammensetzung homogen. Von der Teilnahme ausgeschlossen wurden Personen, die nicht Psychologiestudenten im Grundstudium waren, die keine Erfahrung im Umgang mit Computern hatten oder denen kein E-Mail-Konto zur Verfügung stand. Am zweiten Untersuchungsabschnitt nahmen 80 Studenten anderer Fachrichtungen der Humboldt-Universität zu Berlin teil, welche von den Versuchsleitern direkt am Hauptgebäude der Universität angesprochen wurden.

Material

Den RKs wurden 32 MC-Fragen zum Thema Allgemeinwissen mit beabsichtigten Mängeln [4] vorgelegt. Die Anzahl der eingefügten formalen Fehler variierte zwischen Null und Sechs und wurde mittels Expertenurteil bestätigt. Zwei geschulte Personen beurteilten unabhängig voneinander die Art und Anzahl der Fehler. Im Anschluss wurde in Absprache ein gemeinsames Expertenurteil gebildet. Zur Erleichterung der Gruppenarbeit kam ein Schema zum Einsatz, welches der Bewertung und Verbesserung von MC-Fragen dient und vom Assessment-Bereich der Charité – Universitätsmedizin Berlin entwickelt wurde. Das verwendete Schema ermöglicht die Abfrage zur Einhaltung der Fragenrichtlinien [4] (z.B. un plausible Distraktoren oder Homogenität der Antwortoptionen). Als Maß der Gruppenleistung galt im ersten Untersuchungsabschnitt die Differenz zwischen der Fehlerzahl, welche durch die RKs gefunden wurde und der von Experten bestimmten Fehlerzahl. Im zweiten Untersuchungsab-

schnitt wurde die Leistung über die Erhöhung bzw. Verringerung der Trennschärfen in den Testdurchläufen bestimmt. Die beschriebenen Leistungsmaße dienten der Effektivitätsbestimmung, während die erfasste Bearbeitungszeit als Effizienzkriterium herangezogen wurde. Eine mögliche weitere Effizienzbeurteilung lieferte ein 10 Items umfassender Fragebogen zur Erfassung von Verantwortungsübernahme, Arbeitsweise (bezieht sich auf gemeinsame vs. alleinige Arbeit), Wohlfühl, Motivation und Konzentration auf die Aufgabe. Die verwendeten Items wurden zum Teil gebräuchlichen Fragebögen, wie dem „Fragebogen zur Arbeit im Team“ (FAT; Kauffeld, 2004) entnommen oder als one-item-measurement selbst konstruiert. Bei allen handelt es sich um 6-stufige Likert-Items.

Durchführung

Die Umsetzung der Fragestellung erfolgte mittels eines zweifaktoriellen Designs mit vollständiger Messwiederholung mit den Faktoren Art des Review (face-to-face/virtuell) und Moderation (ja/nein). Die Teilnehmer wurden den RKs randomisiert zugeteilt, wobei die Gruppen letztendlich aus zwei bis vier Teilnehmern bestanden. Dies ergab sich aus dem Umstand, dass manche Versuchspersonen nicht erschienen oder am vereinbarten Termin erkrankt waren. Die Zuordnung zu den Versuchsgruppen erfolgte nach dem Zufallsprinzip, je nach Ankommen zu den Schulungsterminen. Die Reihenfolge der zu durchlaufenden Versuchsbedingungen wurde permutiert, um Reihenfolgeeffekte zu verhindern. Vor Versuchsbeginn wurden die Probanden mit den zu verwendenden Materialien vertraut gemacht und bezüglich der Kriterien zur Beurteilung von MC-Aufgaben geschult. Im Anschluss wurden den Teilnehmern die MC-Fragen (pro Bedingung 8) und Instruktionen per E-Mail übermittelt. Die Zuordnung der Fragen zu den Versuchsbedingungen erfolgte per Zufall. Bevor die Gruppenarbeit begann, stand allen Versuchsteilnehmern eine halbstündige Vorbereitungszeit zur Verfügung. Für die face-to-face Bedingungen waren jeweils 30 Minuten zur Bearbeitung der MC-Fragen geplant, für die virtuellen Bedingungen jeweils eine Woche. Die Moderation wurde von den Versuchsleiterinnen übernommen. Unter jeder Bedingung mussten die vorgelegten MC-Fragen mit Hilfe des o.g. Schemas bewertet und gegebenenfalls verbessert werden. Nach jeder Versuchsbedingung war der Fragebogen zu den Aspekten Verantwortungsübernahme, Arbeitsweise, Wohlfühl, Motivation und Konzentration auf die Aufgabe von den Teilnehmern zu beantworten. Aus der Menge der veränderten MC-Fragen wurde pro Versuchsbedingung für jede der 32 Fragen eine Version zufällig gezogen und zu vier Wissenstests zusammengestellt. Jeder Test enthielt jeweils acht Fragen aus jeder Versuchsbedingung. Ein fünfter Test enthielt die 32 Original-Fragen als Vergleichsmaß für die vorgenommenen Verbesserungen. Die so entstandenen fünf Tests wurden weiteren Studenten abschließend vorgelegt. Die Bearbeitung erfolgte in Anwesenheit der Versuchsleiterinnen.

Statistische Analysen

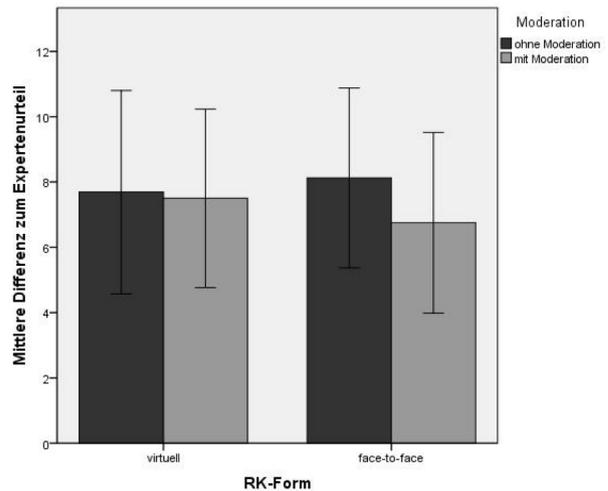
Zur Beurteilung der Gruppenarbeit wurde für die Differenzwerte zwischen Gruppen- und Expertenurteil eine Varianzanalyse (ANOVA) für Messwiederholung berechnet. Zusätzlich zu Testgröße und p-Wert wird jeweils die Effektstärke mit η^2 angegeben. Nach Cohen [5] handelt es sich ab einem η^2 von 0,0099 um einen kleinen Effekt, bei einem η^2 von 0,0588 um einen mittleren und bei $\eta^2=0,1379$ um einen großen Effekt. Die Bearbeitungszeiten der virtuellen Versuchsbedingungen wurden aufgrund der nicht vorliegenden Normalverteilung der Variable ohne Moderation mit Hilfe des Vorzeichenrangtests von Wilcoxon beurteilt. Für die erhobenen Daten zu den Aspekten Verantwortungsübernahme, Arbeitsweise, Motivation, Wohlfühl und Konzentration auf die Aufgabe wurde eine Rangvarianzanalyse für abhängige Stichproben nach Friedman mit den jeweiligen Gruppenmittelwerten durchgeführt. Die im zweiten Teil der Untersuchung erhobenen Trennschärfekoeffizienten der Original-Fragen (punkt-biserialer Korrelationskoeffizient mit Part-Whole-Korrektur) bzw. veränderten Fragen aus den veränderten Bedingungen wurden mittels Wilcoxon-Rangsummen-Test verglichen. Unterschiede galten als statistisch signifikant, wenn $p \leq 0,05$. Alle Daten wurden mit Hilfe des Statistikprogramms SPSS 15.0 analysiert.

Ergebnisse

Effektivität von face-to-face und virtuellen Gruppen

1. Expertenurteil

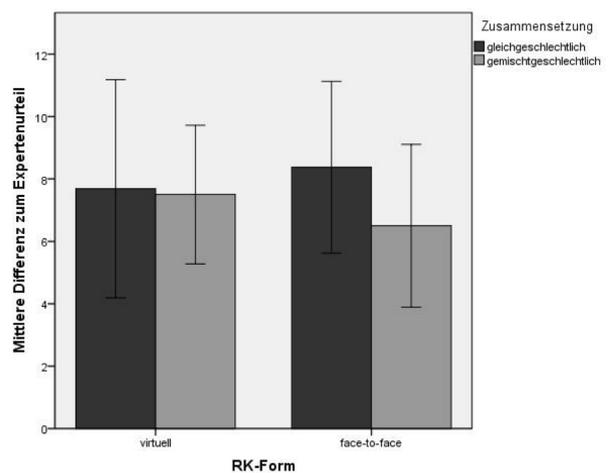
Die Voraussetzungen für die Berechnung der ANOVA wurden geprüft und sind gegeben. In der Bedingung virtuelles Review ohne Moderation beträgt die mittlere Differenz zum Expertenurteil $M=7,69$ ($SD=3,11$) Fehler, in der Bedingung virtuelles Review mit Moderation $M=7,50$ ($SD=2,73$) Fehler. Beim Review face-to-face ohne Moderation werden im Mittel $M=8,12$ ($SD=2,75$) bzw. mit Moderation $M=6,75$ ($SD=2,77$) Fehler der Experten nicht gefunden (siehe Abbildung 1). Weder die RK-Form ($F(1)=0,04$, $p=0,84$, $\eta^2=0,00$) noch der Einsatz von Moderation ($F(1)=1,54$, $p=0,23$, $\eta^2=0,09$) oder die Wechselwirkung der beiden Variablen ($F(1)=0,73$, $p=0,41$, $\eta^2=0,05$) haben einen signifikanten Einfluss auf die Größe der Differenz zum Expertenurteil. Hierbei ist der Anteil der aufgeklärten Varianz für die Variable Moderation mit 9% am höchsten.



Anmerkung: Darstellung der mittleren Differenz zum Expertenurteil als Effektivitätskriterium, Fehlerbalken +/- 1 Standardabweichung

Abbildung 1: Effektivität der RK-Form in Abhängigkeit von Moderation

Die Untersuchung der Wirkung der über alle Versuchsbedingungen konstanten Einflussfaktoren Gruppengröße ($F(2)=0,18$, $p=0,84$, $\eta^2=0,03$) und Geschlechtszusammensetzung ($F(1)=2,06$, $p=0,17$, $\eta^2=0,13$) zeigt ebenfalls keine signifikanten Ergebnisse. In der Versuchsbedingung face-to-face erreichen gleichgeschlechtliche Gruppierungen eine mittlere Differenz zum Expertenurteil von $M=8,38$ ($SD=2,05$) und gemischtgeschlechtliche Gruppierungen $M=6,50$ ($SD=1,04$). Beim virtuellen RK werden in gleichgeschlechtlichen Gruppierungen $M=7,69$ ($SD=3,10$) und in gemischtgeschlechtlichen Gruppierungen $M=7,59$ ($SD=2,40$) Fehler aus dem Expertenurteil gefunden (siehe Abbildung 2).



Anmerkung: Darstellung der mittleren Differenz zum Expertenurteil als Effektivitätskriterium, Fehlerbalken +/- 1 Standardabweichung

Abbildung 2: Effektivität der RK-Form in Abhängigkeit von Zusammensetzung der Gruppen (bzgl. Geschlecht)

2. Reliabilität

Die Reliabilitäten der einzelnen Tests, welche durch die RKs verändert wurden, verbessern sich im Vergleich zu

dem Originaltest (siehe Tabelle 1). Zwei von vier Tests weisen eine deutlich höhere Reliabilität auf.

Tabelle 1: Reliabilitäten des Originaltests und der veränderten Tests

Test	N _{Items} ¹	Reliabilität ²	Reliabilität korr. ³
Original	27	0,268	0,289
Test 1	26	0,590	0,625
Test 2	30	0,676	0,676
Test 3	26	0,343	0,376
Test 4	23	0,299	0,357

Anmerkungen: N_{Items}: Anzahl der Items im jeweiligen Test, ¹unterschiedliche Anzahl von Items aufgrund von Nullvarianz (Schwierigkeit 0 bzw. 100), ²Interne Konsistenz nach Cronbach, ³korrigierte interne Konsistenz nach Cronbach für N_{Items}=30 mit Hilfe der Spearman-Brown-Formel

3. Trennschärfe

Der Wilcoxon-Test zeigt, dass sich die Anzahl von MC-Fragen, deren Trennschärfekoeffizienten nach Bearbeitung durch die RKs höher ausfallen als im Original, von der Anzahl der MC-Fragen, deren Trennschärfekoeffizienten niedriger geworden sind, nur geringfügig unterscheidet. Veränderungen in eine der beiden Richtungen sind bei jeweils etwa der Hälfte der Fragenstichprobe pro Versuchsbedingung zu erkennen. Für die Gesamtheit der Fragen pro Versuchsbedingung kann keine Veränderung der Trennschärfe statistisch nachgewiesen werden (siehe Tabelle 2).

Tabelle 2: Erhöhung bzw. Verringerung der Trennschärfekoeffizienten im Vergleich Original-Test und durch RK veränderte Tests

	Verbesserte Fragen	Verschlechterte Fragen	p
T-Original vs. T-FTFMM	15	12	0,94
T-Original vs. T-FTFOM	15	12	0,74
T-Original vs. T-VMM	14	13	0,87
T-Original vs. T-VOM	13	14	0,26

Anmerkung: T: Trennschärfe; FTFMM: face-to-face mit Moderation; FTFO: face-to-face ohne Moderation; VMM: virtuell mit Moderation; VOM: virtuell ohne Moderation; p: Ergebnis des Wilcoxon-Tests (p-Wert)

Effizienz von face-to-face und virtuellen Gruppen

1. Bearbeitungszeit

Die Bearbeitungszeit beträgt in den beiden face-to-face Bedingungen jeweils konstant 30 Minuten. Für die Gruppenarbeit in der virtuellen Bedingung mit Moderation werden zwischen 5 und 28 Tagen (M=14,31; SD=7,54) benötigt. Währenddessen bearbeiten die Gruppen in der virtuellen Bedingung ohne Moderation ihre Aufgabe in einem Zeitraum zwischen 6 und 55 Tagen (M=12,06; SD=11,69). Hinsichtlich des Effizienzkriteriums Bearbei-

tungszeit unterscheiden sich die verschiedenen virtuellen Bedingungen mit Z=-1,3 und p=0,20 nicht.

2. Weitere Effizienzkriterien

Die durch den Friedman-Test ausgewiesene Rangfolge der Versuchsbedingungen ist für die untersuchten Aspekte Verantwortungsübernahme, Arbeitsweise, Wohlfühl, Motivation und Konzentration auf die Aufgabe gleich (siehe Tabelle 3). Die Bedingung face-to-face ohne Moderation erlangt jeweils die höchsten Werte. Es folgen die Bedingungen face-to-face mit Moderation und virtuell ohne Moderation. Die moderierte virtuelle Bedingung erzielt jeweils den niedrigsten Rangplatz. Die Unterschiede zwischen den Untersuchungsbedingungen sind für alle Aspekte bis auf Motivation statistisch bedeutsam.

Tabelle 3: Friedman-Test zum Vergleich von Verantwortungsübernahme, Arbeitsweise, Wohlfühl, Motivation und Konzentration auf die Aufgabe zwischen den Versuchsbedingungen

Bedingung	V	A	W	M	K
FTFOM	3,34	3,63	3,25	2,66	3,16
FTFMM	2,91	3,13	3,16	2,41	2,72
VOM	2,00	1,84	2,09	2,50	2,28
VMM	1,75	1,41	1,50	2,44	1,84
Chi ²	16,64	33,40	22,40	0,42	10,35
df	3	3	3	3	3
p	0,00**	0,00**	0,00**	0,94	0,02*

Anmerkung: FTFOM: face-to-face ohne Moderation; FTFMM: face-to-face mit Moderation; VOM: virtuell ohne Moderation; VMM: virtuell mit Moderation; df: Freiheitsgrade; p: asymptotische Signifikanz; V: Verantwortungsübernahme; A: Arbeitsweise; W: Wohlfühl; M: Motivation; K: Konzentration auf die Aufgabe; * signifikant bei p < 0,05; ** signifikant bei p < 0,01

Diskussion

Effektivität von face-to-face und virtuellen Gruppen

Zwischen den beiden Gruppen face-to-face und virtuell konnte bzgl. der Entdeckung von Fehlern in MC-Fragen kein Effekt gezeigt werden (0% Varianzaufklärung). Auch hinsichtlich der Erhöhung von Trennschärfekoeffizienten nach dem Review zeigte sich kein Unterschied zwischen den Gruppen. Dieses Ergebnis ist konsistent mit anderen Studien zum Leistungsvergleich [3], [6], [7].

Die Metaanalyse von Baltes [6] konnte keinen Effekt der Gruppengröße identifizieren. Dieser Befund wird durch die Ergebnisse der vorliegenden Studie nur zum Teil gestützt, da wir einen kleinen Effekt zeigen konnten (3% Varianzaufklärung). Der Einflussfaktor Geschlechtszusammensetzung lieferte im Rahmen der ANOVA einen mittleren Effekt (13% Varianzaufklärung). Interessant ist der Fakt, dass die Bearbeitung der Fragen durch ein Review-Komitee, egal welcher Art, die Reliabilität der Tests teilweise deutlich erhöhen konnte.

Effizienz von face-to-face und virtuellen Gruppen

Die vorgegebene Bearbeitungszeit wurde von den face-to-face Gruppen immer eingehalten und die Koordination der Termine erwies sich als unproblematisch. Die Zeitdauer der virtuellen Bedingungen stellte jedoch ein großes Problem in der Zusammenarbeit dar. Zum einen konnte nahezu nie die vorab geplante und mit den Probanden besprochene Dauer von 2 Wochen eingehalten werden und zum anderen entstand dadurch ein erhöhter Koordinationsbedarf durch die Versuchsleiterin.

Die Teilnehmer dieser Untersuchung fühlten sich bei der Zusammenarbeit face-to-face/ unmoderiert wohler als unter allen anderen Arbeitsbedingungen. Sie schätzten sich hier auch als konzentrierter, und verantwortungsvoller ein, und arbeiteten konsequenter als Gruppe zusammen. Computer-medierte Kommunikation ist immer etwas schwerfälliger [8], womit auch der Befund der geringeren Zufriedenheit unter virtuellen Bedingungen erklärt werden kann.

Einfluss der Moderation auf die Gruppenleistung

Die Moderation von Gruppen hat einen mittleren Effekt auf die Leistung der Review-Komitees. Dieses Ergebnis ist konform mit anderen empirischen Untersuchungen zur Leistungssteigerung in Gruppen durch Moderation [3], [9], [10].

Kritische Aspekte

Der zweistufige Schulungsprozess – 3 Stunden für die Versuchsleiter und im Anschluss daran 30 Minuten für die Probanden – könnte zu Informationsverlusten und Verständnisproblemen geführt haben. Zudem verfügen Mitglieder realer RKs über mehr Übung und die regelmäßigen Treffen führen zu einer anderen Gruppendynamik. Das verwendete Schema zur Beurteilung der MC-Fragen bildet im Vergleich mit den Richtlinien von Haladyna [4] nicht vollständig alle Fehlermöglichkeiten ab. Ein weiteres methodisches Problem ist die Anwendung von Messwiederholungen, welche möglicherweise zu Motivationsverlusten führen können. Für die virtuelle Gruppenarbeit konnte nicht kontrolliert werden, wie regelmäßig die Probanden ihr E-Mail-Konto frequentierten und wie zeitnah sie die empfangenen E-Mails beantworteten. Des Weiteren kam es zu unkalkulierbaren technischen Problemen beim Versand der E-Mails, so wurden E-Mail-Adressen von den Probanden falsch angegeben, die empfangenen E-Mails verschwanden im Spam-Ordner oder konnten auf Rechnern mit einem Linux-Betriebssystem nicht geöffnet werden.

Die vollständige Moderationsmethode, deren überlegen effektivitätssteigernde Wirkung verglichen mit Teilaspekten der Methode von Schimansky [10] gezeigt werden konnte, wurde in der vorliegenden Studie nicht zum Ein-

satz gebracht. Eine weitere Limitierung ist der Umstand, dass für diese Studie keine ausgebildeten Moderatorinnen zur Verfügung standen. Aus der Literatur geht hervor, dass professionelle Moderatoren stärker methodisch und aufgabenorientiert vorgehen als solche, die diese Rolle nur "nebenbei" ausfüllen [11].

Da die Motivation grundsätzlich etwas niedriger bewertet wurde als alle anderen Aspekte, ist der geringere Befund unter Umständen gar nicht auf die durchlaufenen Versuchsbedingungen, sondern auf eine geringe initiale Motivation der Studenten zurückzuführen [12]. Ein weiteres Problem der Untersuchung ist die mangelnde Überprüfung der Güte des verwendeten Fragebogens.

Schlussfolgerung

Die elektronische Durchführung des Reviews per E-Mail scheint generell möglich, aber aus Gründen der Effizienz wenig empfehlenswert. Virtuelle Zusammenarbeit ist wesentlich zeitaufwendiger, oftmals mit technischen Problemen verbunden und erfordert einen erhöhten Koordinationsaufwand durch den Moderator. Eine denkbare Alternative könnte die virtuelle Umsetzung nach Nominalgruppentechnik sein. Es ist weitere Forschung an realen RKs notwendig, um beispielsweise differenziertere Leistungsvergleiche zu ermöglichen oder die Rolle des RK-Leiters (Moderator) zu konkretisieren. Vorstellbar könnte auch eine Realisierung virtueller Zusammenarbeit beispielsweise in Internet-Foren sein. Es bleibt zu überprüfen, ob damit eine Eliminierung der oben genannten Probleme gelingen kann.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel haben.

Literatur

1. Scholl W. Grundprobleme der Teamarbeit und ihre Bewältigung - Ein Kausalmodell. In: Gemünden HG, Högl M (Hrsg). Management von Teams. Theoretische Konzepte und empirische Befunde. 3. Auflage. Wiesbaden: Gabler; 2005.
2. Döring N. Sozialpsychologie des Internets. Die Bedeutung des Internets für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen. Internet und Psychologie. Neue Medien in der Psychologie, Band II. Göttingen: Hogrefe; 1999.
3. Unger D, Witte EH. Virtuelle Teams – Geringe Kosten, geringer Nutzen? Zur Leistungsverbesserung von Kleingruppen beim Problemlösen durch elektronische Moderation. Hamburger Forschungsbericht zur Sozialpsychologie (HAFOS), 73. Hamburg: Universität Hamburg, Arbeitsbereich Sozialpsychologie; 2007.
4. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ. 2002;15(3):309-334. DOI: 10.1207/S15324818AME1503_5
5. Cohen J. Statistical power analysis for the behavioral science. 2Nd, edition. Hillsdale, NJ: Lawrence Erlbaum; 1988. S.285.

6. Baltés BB, Dickson MW, Sherman MP, Bauer CC, LaGanke, JS. Computer-mediated communication and group decision making: A meta-analysis. *Organ Behav Hum Decis Process.* 2007;87:156-179. DOI: 10.1006/obhd.2001.2961
7. Li SS. Computer-mediated communication and group decision making. *Small Group Res.* 2007;38(5):593-614. DOI: 10.1177/1046496407304335
8. Becker-Beck U, Wintermantel M, Borg A. Principles of regulating interaction in teams practicing face-to-face communication versus teams practising computer-mediated communication. *Small Group Res.* 2005;36(4):499-536. DOI: 10.1177/1046496405277182
9. Kramer TJ, Fleming GP, Mannis SM. Improving face-to-face brainstorming through modeling and facilitation. *Small Group Res.* 2001;32(5):533-557. DOI: 10.1177/104649640103200502
10. Schimansky A. Die Moderationsmethode als Strukturierungsansatz effektiver Gruppenarbeit. Berlin: Papst; 2007.
11. Immig S, Bachmann T. Was professionelle Moderation leistet. Berlin: artop – Institut an der Humboldt-Universität zu Berlin; 2001. Zugänglich unter/available under: http://www.artop.de/5000_Archiv/5000_PDF_und_Material/Moderation.pdf.
12. Geister S, Konrad U, Hertel G. Effects of process feedback on motivation, satisfaction and performance in virtual Teams. *Small Group Res.* 2006;37(5):459-489. DOI: 10.1177/1046496406292337

Korrespondenzadresse:

Dr. rer. nat. Katrin Schüttpelz-Brauns
Charité - Universitätsmedizin Berlin, Dieter Scheffner
Fachzentrum, Assessment-Bereich / Progress Test
Medizin, Virchowweg 24, D-10117 Berlin, Deutschland
katrin.brauns@charite.de

Bitte zitieren als

Kazubke E, Schüttpelz-Brauns K. Gruppenleistungen beim Review von Multiple-Choice-Fragen - Ein Vergleich von face-to-face und virtuellen Gruppen mit und ohne Moderation. *GMS Z Med Ausbild.* 2010;27(5):Doc68.
DOI: 10.3205/zma000705, URN: urn:nbn:de:0183-zma0007053

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2010-27/zma000705.shtml>

Eingereicht: 09.03.2010

Überarbeitet: 09.07.2010

Angenommen: 20.07.2010

Veröffentlicht: 15.11.2010

Copyright

©2010 Kazubke et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.

Review of multiple-choice-questions and group performance - A comparison of face-to-face and virtual groups with and without facilitation

Abstract

Background: Multiple choice questions (MCQs) are often used in exams of medical education and need careful quality management for example by the application of review committees. This study investigates whether groups communicating virtually by email are similar to face-to-face groups concerning their review process performance and whether a facilitator has positive effects.

Methods: 16 small groups of students were examined, which had to evaluate and correct MCQs under four different conditions. In the second part of the investigation the changed questions were given to a new random sample for the judgement of the item quality.

Results: There was no significant influence of the variables "form of review committee" and "facilitation". However, face-to-face and virtual groups clearly differed in the required treatment times. The test condition "face to face without facilitation" was generally valued most positively concerning taking over responsibility, approach to work, sense of well-being, motivation and concentration on the task.

Discussion: Face-to-face and virtual groups are equally effective in the review of MCQs but differ concerning their efficiency. The application of electronic review seems to be possible but is hardly recommendable because of the long process time and technical problems.

Keywords: multiple choice questions, MCQ, face-to-fac, virtual, facilitation, review-committee

Edda Kazubke¹

Katrin

Schüttpelz-Brauns²

1 Humboldt-Universität zu Berlin, Institut für Psychologie, Berlin, Deutschland

2 Charité - Universitätsmedizin Berlin, Dieter Scheffner Fachzentrum, Assessment-Bereich / Progress Test Medizin, Berlin, Deutschland

Introduction

The regulations for medical licensure require gradings of all courses of the medical training. In many cases multiple choice questions are used in order to test the student's achievements. The quality of the designed questions is of high importance here. To meet this requirement, it is necessary to evaluate the multiple-choice questions and to improve them where applicable before using them in the exams. Usually, the quality control is carried out by review committees (RCs) which are constituted of experts of various faculties and trained to recognize formal errors in questions. This special task is performed in groups to make the experience of many people usable. In order to minimise possible perturbances caused by group dynamic processes, the regular sessions are facilitated. Facilitation has a positive influence on the quality of communication and cooperation [1]. Scheduling the members of the RCs is a problem which could be solved by introducing electronic reviews. Virtual cooperation of the RC could be put into practice via email, for example. The self-determined choice of when to communicate and how long to work on a task is of an advantage [2]. Previously done research

on comparison of work of face-to-face groups and virtual groups could not show differences in the achievements. However, virtual groups needed significantly more time for processing the same task. Moreover, the members of those groups were significantly more unsatisfied with the teamwork. A positive effect of facilitation could also be demonstrated for the field of virtual cooperation [3]. The comparability of the achievements of face-to-face groups and virtual groups as to the special task of reviewing multiple-choice exercises is still extensively unexplored. This leads to the question whether the results of the social psychological research activities can be applied to the area of medical education.

For evaluating the achievements, both effectiveness and efficiency criteria are applied. Furthermore, it is to be investigated to what extent face-to-face groups differ from virtual RC groups with respect to subjective rating of taking over responsibility, teamwork, sense of well-being, motivation and concentration on the task.

Methods

The study was conducted in two stages comprising RCs for evaluating and improving insufficient MCQs and testing the modified questions in order to determine their quality.

Sample survey

41 female students and 8 male students of psychology of the 1st and 2nd year from the Humboldt University in Berlin participated in the RCs, having an average age of 23.17 years (SD=4.21). They were recruited via notices and by individual approach in lectures and credited with test person hours by the Chair in Social Psychology. They worked in 16 small groups of two to four persons. The number of the groups was determined by means of G*Power to evidence an average effect with the help of the chosen testing statistics. 50% of the groups were homogeneously composed with regard to their sex. Excluded from the study were persons who were not students of psychology of the 1st and 2nd year, who were unskilled in using computers or who had no email account at their disposal.

In the second part of the investigation 80 students of other faculties from the Humboldt University in Berlin who were directly approached by the investigators on the university campus participated in the study.

Material

32 MCQs regarding general knowledge containing intended defects [4] were presented to the RCs. The number of inserted formal defects varied from zero to six and was confirmed by an expert's opinion. Two trained persons rated independently from one another the quality and number of the errors. Subsequently, a joint expert opinion was formed by discussion. In order to make the teamwork easier a scheme was used which serves to evaluate and improve multiple-choice questions and which was developed by the assessment center of the Charité - Universitätsmedizin Berlin. The applied scheme allows interrogating compliance with questionnaire guidelines [4] (e.g. implausible distractors or homogeneity of the answer options). During stage one of the study, the group's achievement was measured via the difference between the number of errors which were detected by the RCs and the number of errors determined by the experts. In the second part of the investigation the achievement was determined by increasing and decreasing coefficients of item discrimination in the testing sessions. The described achievement measures were used to determine effectiveness while the recorded processing time served as efficiency criteria. Efficiency rating was further possible by a questionnaire comprising 10 items for recording the aspects taking over responsibility, approach to work (referring to teamwork vs. working on his/her own), sense of well-being, motivation and concentration on the task. The used items were partially taken from usual questionnaires like the German "Fragebogen zur Arbeit im Team"

(FAT, Kauffeld, 2004) (Questionnaire on working in a team), or designed as one-item-measurements. All items are six-point Likert items.

Procedure

The research question was investigated by means of a two-way factorial design with repeated measures of the variables form of review committee (face-to-face/virtual) and facilitation (yes/no). The participants were randomly assigned to the RCs and the groups finally consisted of two to four participants. This was due to the fact that some testing persons failed to appear or had fallen ill on the fixed date. The assignment to the groups was random, depending on the arrival of the participants at the training. The sequence of the testing conditions to be passed was permuted to prevent sequence effects. Prior to the testing, the test persons were familiarized with the material to use and they were trained with regard to the criteria for judging multiple-choice tasks. Then, the MCQs (8 per condition) and instructions were submitted to the participants via email. The assignment of the questions to the testing conditions was carried out at random. Before the teamwork started, the participants had half an hour for preparation. For the face-to-face conditions half an hour was scheduled for answering the MCQs, for the virtual conditions, one week was scheduled. Facilitation was provided by the investigators. In each condition the multiple-choice questions were to be evaluated and, where applicable, improved by means of the above-mentioned scheme. After each test condition the questionnaire was to be answered by the participants as to the aspects taking over responsibility, approach to work, sense of well-being, motivation and concentration on the task. From the set of modified multiple-choice questions one version was randomly selected of each of the 32 questions and put together in four knowledge tests. Each test contained eight questions from each test condition. A fifth test contained the 32 original questions as comparative measure of the improvements. Finally, the five tests generated this way were presented to other students. They were processed in presence of the investigators.

Statistical analyses

The teamwork was evaluated by means of an analysis of variance (ANOVA) for repeated measurements which was calculated for the difference values between the group's and expert's opinion. Additionally to the test statistic and the p-value the effect size is indicated as η^2 . According to Cohen [5], an effect size from $\eta^2=0.0099$ is a small effect, from $\eta^2=0.0588$ is a medium effect and from $\eta^2=0.1379$ is a large effect. The processing times of the virtual test conditions were evaluated by means of the Wilcoxon signed-rank test due to the non-existing normal distribution of the variable without facilitation. With the data collected for the aspects taking over responsibility, approach to work, motivation, sense of well-being, and concentration on the task, an analysis of variance by

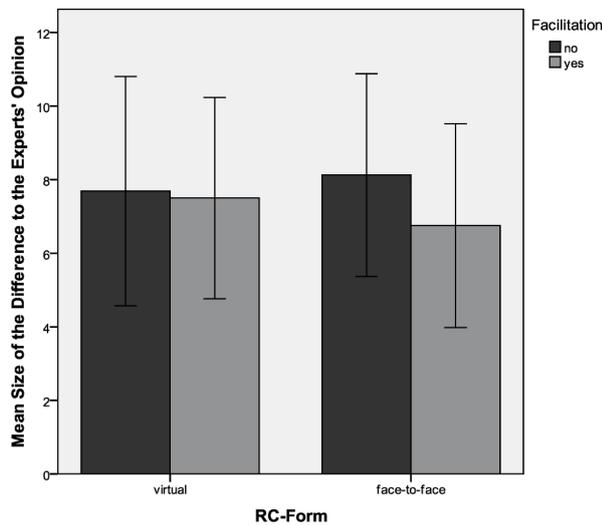
ranks for dependent samples according to Friedman was computed with the respective group means. The coefficients of item discrimination of the original questions collected in the second part of the investigation (point-biserial correlation coefficient with part-whole-correction) and of the modified questions collected in the modified test conditions were compared by means of the Wilcoxon rank-sum test. Differences were considered to be statistically significant with $p \leq 0.05$. All data were analyzed with the SPSS statistics software 15.0.

Results

Effectiveness of face-to-face and virtual groups

1. Expert's opinion

The prerequisites for computing the ANOVA were verified and are given. In the test condition virtual review without facilitation the mean difference to the expert's opinion is $M=7.69$ ($SD=3.11$) errors, in the test condition virtual review with facilitation $M=7.50$ ($SD=2.73$) errors. In the condition review face-to-face without facilitation experts' errors are not detected on an average $M=8.12$ ($SD=2.75$), and with facilitation $M=6.75$ ($SD=2.77$) (see figure 1). Neither the RC form ($F(1)=0.04$, $p=0.84$, $\eta^2=0.00$) nor applying facilitation ($F(1)=1.54$, $p=0.23$, $\eta^2=0.09$) or interaction of both variables ($F(1)=0.73$, $p=0.41$, $\eta^2=0.05$) have significant influence on the size of the difference to the expert's opinion. In this case, the shared explained variation for the variable facilitation is highest with 9%.

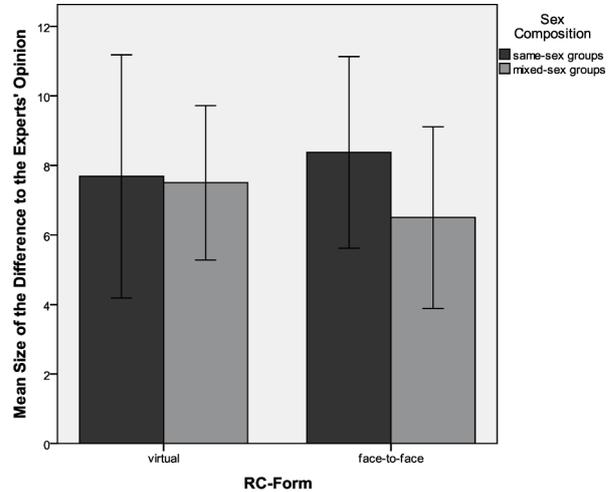


Note: Representation of the mean difference to the expert's opinion as effectiveness criteria, error bar +/- 1 standard deviation

Figure 1: Effectiveness of the RC form depending on facilitation

The investigation of the effect of the influencing factors group size ($F(2)=0.18$, $p=0.84$, $\eta^2=0.03$) and sex composition ($F(1)=2.06$, $p=0.17$, $\eta^2=0.13$) which were constant over all test conditions does not show any significant results either. In the test condition face-to-face the mean

difference to the expert's opinion of same-sex groups is $M=8.38$ ($SD=2.05$) and of mixed-sex groups is $M=6.50$ ($SD=1.04$). In the virtual RC errors of the expert's opinion are detected in same-sex groups with $M=7.69$ ($SD=3.10$) and in mixed-sex groups with $M=7.59$ ($SD=2.40$) (see figure 2).



Note: Representation of the mean difference to the expert's opinion as effectiveness criteria, error bar +/- 1 standard deviation

Figure 2: Effectiveness of the RC form depending on group composition (referring to sex)

2. Reliability

The reliabilities of the respective tests which were modified by the RCs improve in comparison to the original test (see table 1). Two of four tests show relevant higher reliability.

Table 1: Reliabilities of the original test and the modified tests

Test	N_{Items}^1	Reliability ²	Reliability corr. ³
Original	27	0.268	0.289
Test 1	26	0.590	0.625
Test 2	30	0.676	0.676
Test 3	26	0.343	0.376
Test4	23	0.299	0.357

Notes: N_{Items} : Number of items in the respective test, ¹ different number of items due to zero-variance (difficulty 0 and 100), ² internal consistency according to Cronbach's alpha, ³ corrected internal consistency according to Cronbach's alpha for $N_{\text{Items}}=30$ by means of the Spearman-Brown formula

3. Item discrimination

The Wilcoxon test shows that the number of multiple-choice questions whose coefficient of discrimination is higher than originally after modification by the RCs differs only slightly from the number of multiple-choice questions whose coefficient of discrimination decreased. Changes into one of the two directions can be recognized in approximately 50% of the question sample per test condition.

For the overall of the questions per test condition a change in the power of the test cannot be proved statistically (see table 2).

Table 2: Increase and decrease of the item discrimination coefficients comparing the original test and tests modified by RC.

questions	Improved questions	Deteriorated	p
D-Original vs. D-FTFMM	15	12	0.943
D-Original vs. D-FTFOM	15	12	0.737
D-Original vs. D-VMM	14	13	0.866
D-Original vs. D-VOM	13	14	0.264

Note. D: item discrimination; FTFMM: face-to-face with facilitation; FTFOM: face-to-face without facilitation; VMM: virtual with facilitation; VOM: virtual without facilitation; p: result of the Wilcoxon test (p-value)

Efficiency of face-to-face and virtual groups

1. Processing time

In both face-to-face conditions the processing time is constantly 30 minutes. The teamwork in the virtual test condition with facilitation requires between 5 and 28 days ($M=14.31$; $SD=7.54$). Meanwhile, the groups in the virtual test condition without facilitation process the task within 6 and 55 days ($M=12.06$; $SD=11.69$). As to the efficiency criteria processing time, the various virtual conditions do not differ with $Z=-1.3$ and $p=0.20$.

2. Further efficiency criteria

The ranking of the test conditions shown by the Friedman test is identical for the investigated aspects taking over responsibility, approach to work, sense of well-being, motivation and concentration on the task (see table 3). The condition face-to-face without facilitation reaches the highest scores in each case. This is followed by the conditions face-to-face with facilitation and virtual without facilitation. The virtual facilitated condition reaches the lowest score in each case. The differences between the test conditions are statistically relevant for all aspects except motivation.

Table 3: Friedman test for comparison of taking over responsibility, approach to work, sense of well-being, motivation and concentration on the task between the test conditions

Condition	R	A	W	M	C
FTFOM	3.34	3.63	3.25	2.66	3.16
FTFMM	2.91	3.13	3.16	2.41	2.72
VOM	2.00	1.84	2.09	2.50	2.28
VMM	1.75	1.41	1.50	2.44	1.84
Chi ²	16.64	33.40	22.40	0.42	10.35
df	3	3	3	3	3
p	0.00**	0.00**	0.00**	0.94	0.02*

Note: FTFOM: face-to-face without facilitation; FTFMM: face-to-face with facilitation; VOM: virtual without facilitation; VMM: virtual with facilitation; df: degrees of freedom; p: asymptotic significance; R: taking over responsibility; A: approach to work; W: sense of well-being; M: motivation; C: concentration on the task; * significant at $p < 0.05$; ** significant at $p < 0.01$

Discussion

Effectiveness of face-to-face and virtual groups

As to detecting errors in MCQs no effect between the face-to-face and the virtual groups could be proved (0% explained variation) Also, as far as the increase of discriminatory power coefficients after the review is concerned, no difference between the groups was found. This result is consistent with other achievement comparison studies [3], [6], [7].

The Baltes meta-analysis [6] could not identify any effect of the group size. This result is only partly supported by the findings of the present study since we were able to prove a small effect (3% explained variation). The influencing factor sex composition showed a medium effect within the framework of the ANOVA (13% explained variation). Interesting is the fact that the processing of the questions by a review committee, no matter of which kind, could increase the reliability of the tests in parts relevantly.

Efficiency of face-to-face and virtual groups

The allowed processing time limit was always met by the face-to-face groups and coordinating the fixed dates was uncomplicated. Yet, the period of time in the virtual conditions turned out to be a big problem for the teamwork. On the one hand, the time limit of two weeks fixed with the probands was nearly never met, on the other hand,

this produced an increased demand for coordination by the investigator.

The participants of this investigation felt more comfortable with face-to-face/unfacilitated cooperation than in the other working conditions. In addition, they rated themselves as more concentrated and responsible and worked more consistently as a group together. Computer-mediated communication is always a bit cumbersome [8], which explains the finding of the lower satisfaction in virtual conditions.

Impact of facilitation on the group's achievement

Facilitating groups has a medium effect on the achievements of the review committees. This result is conform with other empirical investigations on increasing achievement in groups via facilitation [3], [9], [10].

Critical aspects

The two-stage training process (3 hours for the investigators followed by 30 minutes for the test persons) might have led to loss of information and problems in understanding. In addition, members of real RCs have more practice and the regular meetings lead to a different kind of group dynamic. The scheme used for judging the multiple-choice questions does not reproduce completely all possibilities of defects in comparison to the guidelines of Haladyna [4]. Another methodical problem is the use of repeated measurements which might lead to loss in motivation. With regard to the virtual teamwork it could not be controlled how often the test persons checked their email accounts and how fast they answered the received emails. Furthermore, unforeseeable technical problems occurred when sending the emails: Email addresses had been misstated by the test persons, the received emails disappeared in spam folders or could not be opened on computers with a Linux operating system. The complete facilitation method which has, as could be demonstrated by Schimansky [10], a superior impact on increasing effectiveness in comparison to partial aspects of the method, was not used in the present study. Another limiting factor is that no trained facilitators were available for this investigation. As can be seen from the literature, professional facilitators proceed more methodically and oriented on tasks than those who perform this role solely casually [11].

Since motivation was basically rated lower than all other aspects, the lower score might not trace back to the passed test conditions but to a low initial motivation of the students [12]. Another problem of the investigation is insufficient examination of the quality of the used questionnaire.

Conclusion

The application of electronic review via email seems to be possible but is hardly recommended due to efficiency reasons. Virtual teamwork is considerably more time consuming, often associated with technical problems and requires a higher demand for coordination by the facilitator. A possible alternative could be a virtual application according to nominal group technique. Further investigation on real RCs is necessary to permit for example more differentiated achievement comparisons or to substantiate the role of the RC leader (facilitator). For example, the realization of virtual teamwork in internet forums could be imagined. It remains to be verified whether eliminating the above problems is possible in this way.

Competing interests

The authors declare that they have no competing interests.

References

1. Scholl W. Grundprobleme der Teamarbeit und ihre Bewältigung - Ein Kausalmodell. In: Gemünden HG, Högl M (Hrsg). Management von Teams. Theoretische Konzepte und empirische Befunde. 3. Auflage. Wiesbaden: Gabler; 2005.
2. Döring N. Sozialpsychologie des Internets. Die Bedeutung des Internets für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen. Internet und Psychologie. Neue Medien in der Psychologie, Band II. Göttingen: Hogrefe; 1999.
3. Unger D, Witte EH. Virtuelle Teams - Geringe Kosten, geringer Nutzen? Zur Leistungsverbesserung von Kleingruppen beim Problemlösen durch elektronische Moderation. Hamburger Forschungsbericht zur Sozialpsychologie (HAFOS), 73. Hamburg: Universität Hamburg, Arbeitsbereich Sozialpsychologie; 2007.
4. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ.* 2002;15(3):309-334. DOI: 10.1207/S15324818AME1503_5
5. Cohen J. Statistical power analysis for the behavioral science. 2Nd, edition. Hillsdale, NJ: Lawrence Erlbaum; 1988. S.285.
6. Baltés BB, Dickson MW, Sherman MP, Bauer CC, LaGanke, JS. Computer-mediated communication and group decision making: A meta-analysis. *Organ Behav Hum Decis Process.* 2007;87:156-179. DOI: 10.1006/obhd.2001.2961
7. Li SS. Computer-mediated communication and group decision making. *Small Group Res.* 2007;38(5):593-614. DOI: 10.1177/1046496407304335
8. Becker-Beck U, Wintermantel M, Borg A. Principles of regulating interaction in teams practicing face-to-face communication versus teams practising computer-mediated communication. *Small Group Res.* 2005;36(4):499-536. DOI: 10.1177/1046496405277182
9. Kramer TJ, Fleming GP, Mannis SM. Improving face-to-face brainstorming through modeling and facilitation. *Small Group Res.* 2001;32(5):533-557. DOI: 10.1177/104649640103200502

10. Schimansky A. Die Moderationsmethode als Strukturierungsansatz effektiver Gruppenarbeit. Berlin: Papst; 2007.
11. Immig S, Bachmann T. Was professionelle Moderation leistet. Berlin: artop – Institut an der Humboldt-Universität zu Berlin; 2001. Zugänglich unter/available under: http://www.artop.de/5000_Archiv/5000_PDF_und_Material/Moderation.pdf.
12. Geister S, Konrad U, Hertel G. Effects of process feedback on motivation, satisfaction and performance in virtual Teams. Small Group Res. 2006;37(5):459-489. DOI: 10.1177/1046496406292337

Please cite as

Kazubke E, Schüttpelz-Brauns K. Gruppenleistungen beim Review von Multiple-Choice-Fragen - Ein Vergleich von face-to-face und virtuellen Gruppen mit und ohne Moderation. GMS Z Med Ausbild. 2010;27(5):Doc68.
DOI: 10.3205/zma000705, URN: urn:nbn:de:0183-zma0007053

This article is freely available from

<http://www.egms.de/en/journals/zma/2010-27/zma000705.shtml>

Received: 2010-03-09

Revised: 2010-07-09

Accepted: 2010-07-20

Published: 2010-11-15

Corresponding author:

Dr. rer. nat. Katrin Schüttpelz-Brauns
Charité - Universitätsmedizin Berlin, Dieter Scheffner
Fachzentrum, Assessment-Bereich / Progress Test
Medizin, Virchowweg 24, D-10117 Berlin, Deutschland
katrin.brauns@charite.de

Copyright

©2010 Kazubke et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share – to copy, distribute and transmit the work, provided the original author and source are credited.