

Multiple- Choice-Tests (MC-Tests)

MC-Tests zielfördernd formulieren –
ein Leitfaden für Lehrpersonen

Verfasser/in: Ambroise Baillifard und Alice Thurre-Millius
Datum: April 2020
Zeitaufwand für die Lektüre: 30 Minuten

Zusammenfassung

Inhaltsverzeichnis

1	Einleitung	2
	2	
2	Die vier Ausrichtungen von MC-Tests	4
	4	
2.1	Diagnose: Anregung durch MC-Tests	5
2.2	Entdecken: «Formative» MC-Fragen zur Förderung von Einsatz und Motivation	6
2.3	(Tutorielle) Begleitung: MC-Tests mit «Bildungseffekt» zur Förderung des Lernens	6
2.4	Beurteilungen : «Summative» MC-Test (Validität, Hierarchisierung und Benotung)	8
3	Ausarbeitung	12
3.1	Die Anleitung	12
3.2	Die Problemstellung	12
3.3	Die korrekte(n) Antwort(en) (bzw. Antwortmöglichkeiten)	13
3.4	Die Distraktoren (Fallen/Fangfragen, falsche Antwortmöglichkeiten)	13
3.5	Formative Rückmeldung (Feedback)	14
4	Vorgefasste Meinungen	16
4.1	Vorgefasste Meinung Nr. 1: Treffer können zufallsbedingt sein	16
4.2	Vorgefasste Meinung Nr. 2: Die korrekte Antwort lässt sich durch Ausschluss der Distraktoren ableiten.	17
4.3	Vorgefasste Meinung Nr. 3: GMC-Tests zu entwerfen ist zeitaufwändig	18
4.4	Vorgefasste Meinung Nr. 4: Mit MC-Tests lässt sich nicht messen, ob die Studierenden in der Lage sind, analytisch zu denken oder vom Stoff zu abstrahieren.	18
5	Zusammenfassung: Fallstricke und Hinweise	22
6	Vorzüge und Grenzen	24
7	Bibliographie	25
7.1	Wissenschaftliche oder pädagogische Werke	25
7.2	Ergänzende Bibliographie	26
7.3	Weitere Werke	26

1 Einleitung

Wieso zählen bestimmte Tests (z. B. Mini-Quiz, MC-Fragen, ...) zu den erfolgreichsten Lernmitteln?

- a) Diese Aussage ist falsch.
- b) Sie motivieren die Studierenden, die Antwort aus dem Gedächtnis abzurufen.
- c) Sie unterstützen den Lernerfolg, indem sie erworbenes Wissen neu strukturieren.
- d) Sie sorgen dafür, dass die Studierenden sich an Antworten herantasten.
- e) Sie bewirken, dass die Studierenden länger arbeiten.

Richtig: b) und c)

Die Lehrperson stellt eine Frage, der/die Studierende antwortet; die Lehrperson stimmt zu oder nicht, sie nimmt Fortschritte oder Wissenslücken zur Kenntnis. Für jede gewissenhafte Lehrperson stellt die Evaluation des Wissensstands ein Kernproblem dar.

- Wie lässt sich Wissen messen und bewerten? Wie lässt sich sicherstellen, dass dasjenige Wissen evaluiert wird, das effektiv evaluiert werden soll?
- Welche Art von Evaluation ist angebracht?
- Wie kann eine Lehrperson überprüfen, dass die von ihr erstellte Evaluation tatsächlich das misst, was die Lehrperson messen möchte?
- Wie hoch sollen die Anforderungen für das Bestehen des Tests (Bestehensgrenze) sein?
- Wie lässt sich überprüfen, ob der/die Studierende fähig ist, sein/ihr Wissen anzuwenden?
- Etc.

Diese Anleitung befasst sich mit einer einzigen Art der Evaluation: dem MC-Test in Fragebogenform. Wir beschränken uns dabei nicht auf den Einsatz von MC-Tests als Evaluationsinstrument, sondern behandeln das Konzept MC-Fragen im weiteren Sinn, da sie sich während des ganzen Lernprozesses einsetzen lassen. Brown et al. (2014) werten Mini-Quiz als wesentlich effizientere Hilfsmittel zur Wiederholung als das erneute Durchlesen des Stoffs. «Als Auslösereiz für Gedächtnisanstrengungen verstärken sie die Lerneffekte und unterbrechen den Vergessensprozess. Ein einziges, einfaches Quiz in Anschluss an die Lektüre eines Texts oder am Ende einer Lektion ist ein nützlicheres Instrument zum Erwerb von Wissen als das erneute Durchlesen des Texts oder der eigenen Notizen.» (S. 23) [Übersetzung].

Sie sehen, MC-Fragen – eine gleichermassen verteilte und weit verbreitete Lernhilfe (Bertrand, 2014) – eignen sich nicht nur für Evaluationen, sondern auch zur Vermittlung von Stoff, zum Unterricht, um das Interesse zu wecken, das Lernen zu erleichtern, Anregungen zu geben und zu motivieren. Im Folgenden möchten wir aufzeigen, wie sich MC-Fragen einsetzen lassen. Dieses Dokument versteht sich als Leitfaden für Lehrpersonen: Es soll keinen alleinseligmachenden Weg vorgeben, sondern Ansätze vermitteln und insbesondere dazu anregen, MC-Tests als Lerninstrument zu nutzen. Drei Leitgedanken bilden den Kern unserer Ausführungen:

Leitgedanken	In welchem Kapitel wird auf sie genauer eingegangen?
Die höheren Kognitionsebenen (Analyse, Synthese, Urteilsvermögen) lassen sich anhand von MC-Fragen testen. Derzeit dienen MC-Fragen allerdings zumeist zur Messung des Wissensstandes.	4.4 Vorgefasste Meinung Nr. 4
Rückmeldungen (Feedbacks) tragen nicht nur zum Lernerfolg, sondern auch zur Motivation bei und wecken Begeisterung. Feedbacks können befreiend wirken: einerseits als Befreiung von der Illusion, etwas zu wissen, andererseits als Befreiung von der Angst, Fehler zu machen und nicht zu begreifen, weshalb. Dank Feedbacks erwerben die Studierenden die Fähigkeit, sich selber sachkundig zu beobachten.	3.5 Formative Rückmeldung (Feedback)
Ein MC-Test lässt sich zur summativen Beurteilung einsetzen, sofern er sorgfältig aufgebaut ist und mathematisch kohärente <i>Scoring</i> -Regeln einhält.	2.4 Beurteilungen: Validität und Hierarchisierung bei «summativen» MC-Tests

2. Die vier Ausrichtungen von MC-Tests

Was bezwecken MC-Tests? Je nach Ausrichtung werden diese Fragebögen unterschiedlich aufgebaut. Es wird zwischen vier verschiedenen Ausrichtungen unterschieden:

1. **Diagnose:** Neugierde wecken, (neue) Themen einführen, Ausgangskennntnisse ermitteln, Aufzeigen, dass der gesunde Menschenverstand nicht immer ausreicht und so Interesse an bestimmten Fragen wecken;
2. **Erkenntnis:** Die Studierenden sollen erkennen, dass bestimmte Konzepte erlernt werden müssen. Nicht selten merken Studierende erst an der Prüfung, dass sie den Stoff nicht verstanden haben.
3. **Begleitung:** Feedbacks erlauben es, den Lernprozess zu optimieren. Studierende erarbeiten Lernstoff effizienter, wenn sie Rückmeldungen zu ihren Fehlern erhalten.
4. **Bewertung:** Bewertungen sind möglich, wenn sich die Frage auf ein Lernziel bezieht und anlässlich einer Prüfung gestellt wird.

Je nach Kontext, lässt sich ein und dieselbe Frage mit verschiedenen Ausrichtungen stellen: Sie kann in unterschiedlichen Lernphasen eingesetzt werden. Die beiden folgenden Fragen lassen sich beispielsweise verwenden, um Neugierde zu wecken, den Wissensstand der Studierenden zu ermitteln, sie tutoriell zu begleiten oder ihre Kenntnisse zu bewerten.

Wieviel wiegt eine Wolke?

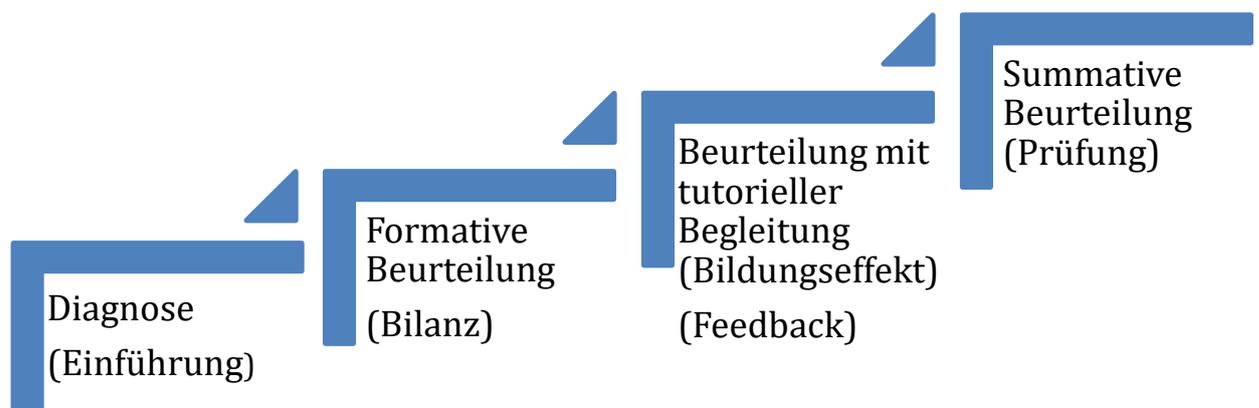
- a) Jede Wolke wiegt rund eine Tonne.
- b) Eine kleine Wolke wiegt rund 25 Tonnen.
- c) Eine kleine Wolke wiegt rund 1000 Tonnen.
- d) Wolken haben kein Gewicht, auch wenn sie gross sind.
- e) Das kann man nicht wissen, Wolken wurden noch nie gewogen.

Sie finden das Feedback auf Seite 5.

Weswegen ist es am Südpol kälter als am Nordpol?

- a) Wegen der Gletscher
- b) Wegen der Eisberge
- c) Wegen der Höhenlage
- d) Wegen des Breitengrades
- e) Wegen der Neigung der Erdachse

Sie finden das Feedback auf Seite 5.



2.1 Diagnose: Anregung durch MC-Tests

Der Einführung von Lerninhalten oder dem Erwerb von Wissen vorgelagerte Tests sind effektiv sehr nützlich, sofern Irrtümer im Anschluss durch Rückmeldungen ausgeräumt (korrigiert) werden.

(Brown, Roediger und McDaniel, 2014)

Ein solcher Test kommt zu Beginn der Bildungseinheit zum Einsatz, um den vorhandenen Wissensstand zu ermitteln oder ein neues Thema einzuführen und herauszufinden, was bei den Studierenden bereits bekannt ist. Diese Verwendung ist selten, kann sich aber in folgenden Fällen als nützlich erweisen:

- Um eine Entwicklung aufzuzeigen: im Anschluss an Präsenzunterricht, vor Beginn einer Lernperiode. Planen Sie beispielsweise für einen diagnostischen MC-Test einen Einmaldurchgang ein und legen Sie ihn den Studierenden nach drei oder vier Wochen erneut vor. So werden Fortschritte sichtbar und die Motivation steigt. Moodle ist derart konfiguriert, dass Sie Ihren MC-Test während einer vorab festgelegten Dauer oder als zwei Einzeldurchläufe mit vorab festgelegtem Intervall anzeigen lassen können.
- Mithilfe von MC-Tests können Sie Ihren Unterricht auch anpassen, falls Ihren Studierenden wichtige Grundlagen fehlen oder – im Gegenteil – Teile des Unterrichtsinhalts bereits bekannt sind.
- Mit MC-Tests können Sie das Interesse der Studierenden wecken / sie auf zweckmässige Elemente ausrichten, wenn sie sich mit dem Unterrichtsmaterial befassen.
- MC-Tests (z. B. in Form von Umfragen) tragen dazu bei, dass sich Lehrpersonen im Präsenz- oder virtuellen Unterricht an die Studierenden anpassen können.

Es ist seltsam, aber die diagnostischen MC-Tests haben bis heute keine gebührende Beachtung gefunden. Zeit, dies zu ändern!

Sonderfall: Rhetorische Fragen

Die Verfasser der «Encyclopédie qui raconte» (1996) haben eine Reihe von Fragen formuliert, bei denen die jugendlichen Lesenden die Antworten nicht kennen können. Hier einige Beispiele:

- Warum hat ein Nordamerikanischer Ochsenfrosch drei Augenlider?
- Wie finden Brieftauben wieder nach Hause?
- Was frisst das Schnabeltier?
- Weshalb reiben Fliegen ihre Beine aneinander?
- Raupen fressen Blätter. Schmetterlinge auch?
- ...

Die Verfasser gehen nicht davon aus, dass das Publikum diese Fragen beantworten kann. Die Fragen sollen vielmehr die Neugierde wecken sowie das Lernen und die Suche nach Informationen fördern.

Diese Art von Fragen zeigen den Lesenden einerseits Wissenslücken auf und motivieren sie andererseits, den anschliessenden Text aufmerksam zu lesen. Diese Art von Fragen lassen sich gut in Handouts einfügen, um die einzelnen Kapitel einzuführen.

2.2 Entdecken: «Formative» MC-Fragen zur Förderung von Einsatz und Motivation

Paradoxerweise führt guter Unterricht dazu, dass die Studierenden in der Illusion von Wissen leben. Tendenziell merken diese Studierenden nicht, dass sie den Stoff nicht beherrschen. Der Grund liegt darin, dass die Handouts und die Aussagen der Lehrpersonen sehr klar gestaltet sind. Studierende, die im Unterricht «gut mitkommen», glauben rasch, dass sie auch den Stoff beherrschen. Wer kennt diese Situation nicht? Wenn man Sie auffordert, das im Unterricht Gelernte mit eigenen Worten wiederzugeben, sind Sie dazu nicht imstande. Ein äusserst unangenehmes Gefühl ... Vor allem wenn Sie dachten, dass Sie alles verstanden hätten. Formative MC-Tests ermöglichen es, im Verlauf des Moduls bzw. des Semesters eine Zwischenbilanz zu ziehen. Mit diesen Tests lässt sich prüfen, ob die Studierenden effektiv Wissen erworben haben. Sie liefern Anhaltspunkte für die Benotung, die bei der Bewertung der Unterrichtseinheit nicht in Anrechnung gebracht werden, und zeigen die Fortschritte der Studierenden auf. Sie können autonom (Selbstevaluation) oder von der Lehrperson (formative Evaluation) überprüft werden.

Formative MC-Tests bieten somit eine Möglichkeit zu Zwischenbilanzen. Sie bringen die Studierenden dazu, «gefühltes Wissen» nicht länger als Wissen einzustufen, und zeigen ihnen den Lernbedarf auf. Formative MC-Tests sind daher ein zweischneidiges Schwert: Sie können ein Auslöser sein, um sich erneut mit dem Stoff zu beschäftigen, sie können aber auch entmutigen. Daher dürfte es sich lohnen, wenn Sie den Schwierigkeitsgrad der einzelnen MC-Tests insgesamt anpassen und darauf achten, dass er weder zu leicht noch zu entmutigend ausfällt. Folgendes sollten wir dabei nicht vergessen: «Übermässige Strenge [...] entmutigt die Studierenden; sie verzweifeln und quälen sich. Ausserdem – dies ist im Übrigen das Schädlichste – entwickeln sie Angst vor allem und trauen sich daher gar nichts mehr. [...] Die Hoffnung ist die stärkste Lernmotivation» (Quintilian, II 10–13) [Übersetzung].

Dieser Diskurs nimmt allerdings eine ganz andere Richtung, wenn Sie die Perspektive wechseln. Wenn Sie die Tiefe des erworbenen Wissens anstatt der Motivation als Kriterium annehmen, sollten die Fragen möglichst komplex sein und möglichst häufig wiederholt werden. Regelmässigkeit und kognitive Anstrengungen, wenn sich die Studierenden den Stoff erneut vor Augen führen wollen, sorgen wirksam für Lerneffekte. Fragen, bei denen man mit mehreren Konzepten gleichzeitig umgehen und Stoff aus den tiefsten Tiefen des Gedächtnisses abrufen muss, führen dazu, dass das Gehirn seine Erinnerungen konsolidiert und sich weitere Wissensabrufe vereinfachen. Tatsächlich verhält es sich so, dass diejenigen Unterrichtsteile am dauerhaftesten erworben werden, die im Verlauf des Semesters in kurzen Abfragen getestet werden. Dies gilt vor allem, wenn die Studierenden korrigierte und kommentierte Antworten erhalten (Brown P.C. et al., 2014).

Als Lehrperson können Sie beispielsweise einen obligatorischen MC-Test mit einer bestimmten Bestehensgrenze als Vorbedingung für die Zulassung zu einer Prüfung oder die Abgabe eines Essays ansetzen. So machen Sie deutlich, dass die Studierenden bestimmte Lerninhalte beherrschen müssen, bevor sie kreativ mit ihnen umgehen können.

2.3 (Tutorielle) Begleitung: MC-Tests mit «Bildungseffekt» zur Förderung des Lernens

Aussagekräftige Fragebogenzusätze in Textform (Hinweise, Tipps, Rückmeldungen, Kommentare, Hilfestellungen) vereinfachen das Üben und das Training, fördern den Fortschritt und erhöhen die Transparenz der Beurteilung. Somit werden formative MC-Tests zu MC-Tests mit Bildungseffekt. Derartige Fragebögen beziehen eine spezifische Form von Paratexten mit ein: **formative Feedbacks**. Wir unterscheiden zwischen formativen MC-Tests und MC-Tests mit Bildungseffekt, um die Wichtigkeit solcher Feedbacks zu betonen (auch als «**formative Rückmeldungen**» bezeichnet). Für Sie als Lehrperson geht es darum, das Lernen durch indikative Hilfestellungen zu erleichtern. Denken Sie beispielsweise an Lösungshilfen, Erläuterungen, eingehende Antworten, Wechsel des Blickwinkels, Betrachtungen des Gegenstands aus einem gewissen Abstand, Weiterführungen ... Derartige Feedbacks erfolgen häufig bei fehlerhaften Antworten; Sie können aber auch positive Rückmeldungen auf korrekte Antworten geben und so den Blickwinkel erweitern.

Lernwirksame Feedbacks (Hilfestellungen) zu den Fragen auf Seite 3:

Lässt es sich überhaupt vermeiden, bestimmte Fragen mit einem Nachsatz zu versehen, um sie subtiler zu gestalten? Die nachstehenden Feedbacks sind wesentlicher als die eigentlichen Fragen. Die Fragen wecken die Neugier der Lesenden und die Feedbacks motivieren sie dazu, mehr herauszufinden. Wenn die Studierenden schon vor ihren Antworten wüssten, worauf die Fragen abzielen oder was hinter ihnen steht, wären die Fragen an sich überflüssig.

a) Gewicht einer Wolke.

Eine kleine Wolke wiegt rund 25 Tonnen. Auch wenn man eine Wolke nicht auf die Waage legen kann, lässt sich ihre Masse doch schätzen. Man schätzt ihr Volumen und damit ihren Wasser- und Eisgehalt (ein Millionstel des Gewichts einer Wolke besteht aus Wasser). Eine kleine Wolke von 20-30 Millionen m³ enthält also ein Wasservolumen von ganzen 20-30 m³.

b) Eis am Südpol

Der Südpol liegt höher als der flach auf dem Packeis liegende Nordpol, an dem immerhin eine Temperatur von minus 30° herrscht. Am Südpol (2800 m. ü. M.), sinkt die Temperatur auf minus 60° ab.

Mithilfe von MC-Tests mit Bildungseffekt stellen Sie mehr Transparenz bezüglich der vor Ihnen gesetzten Lernziele und Ihrer Erwartungen als Lehrperson her, Sie steigern das Interesse am Lernauftrag und am Unterrichtsgegenstand (Sie erweitern ihn, bieten neue Blickwinkel und illustrieren Regeln durch Beispiele) und verringern die Misserfolgs- und Abbrecherquoten. Diese Art von MC-Tests möchten wir den Lehrpersonen ganz besonders ans Herz legen. Die «lernwirksamen» Feedbacks (Beschreibung von Shute, 2008) sollen das Verständnis bei fortgeschrittenen Studierenden fördern, anstatt ihnen vorzuschreiben, was sie zu tun haben. Vermeiden Sie folgende Arten von Feedbacks: «Diese Antwort ist falsch» oder «Siehe Lehrbuch, Seite XXX». Bransford et al. (2000) geben folgende Anregungen:

- Vermitteln Sie Ansätze (Hilfestellungen durch Formulierung der Problemstellung oder lernwirksame Hilfestellung, Hinweise), die die Erledigung des Lernauftrags erleichtern;
- Zeigen Sie auf, worin sich die Antworten der einzelnen Studierenden von den erwarteten Antworten unterscheiden;
- Liefern Sie klare Vorgaben für das, was das Lernen erbringen soll (Bewertung, Notenskala, Anforderungsniveau, Anzahl korrekter Antworten).

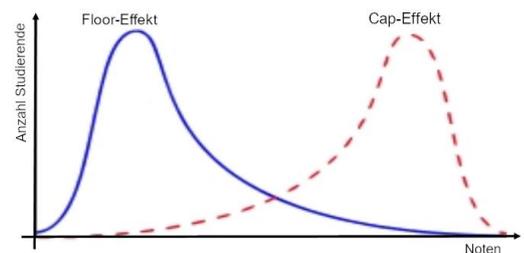
Selbstverständlich sollte ein lernwirksames Feedback nicht so gestaltet werden, dass es sich einfach ab-schreiben und als Antwort einsetzen lässt. Lernwirksames Feedback erfolgt, nachdem die Studierenden ihre Antworten abgegeben haben. Einige Tipps für Lehrpersonen:

- Sehen Sie eine Gedächtnisstütze («Eselsbrücke») für Details vor;
- Begründen Sie, weshalb eine bestimmte Antwort korrekt ist;
- Machen Sie Angaben zum Unterrichtsinhalt, auf den sich eine spezifische Frage bezieht;
- Zeigen Sie den Lösungsweg auf, ohne die Frage zu beantworten;
- Zeigen Sie die Fehler auf und begründen Sie, weshalb es sich um Fehler handelt: geben Sie Informationen zu den Fehlern;
- Geben Sie strategische Hinweise zum Verfahren: «Um den Hochpunkt dieser Funktion f zu finden, ist die Ableitung zu verwenden. Die Ableitung einer Funktion entspricht ihrer Neigung. Beträgt die Ableitung von f Null, so ist die Neigung *horizontal*. Ermitteln Sie nun, ob es sich hier um einen Hoch- oder einen Tiefpunkt handelt.»
- Geben Sie ein Fallbeispiel, das mit der Frage in Zusammenhang steht;
- Verweisen Sie auf eine wissenschaftliche Debatte zum Thema.

2.4 Beurteilungen: «Summative» MC-Test (Validität, Hierarchisierung und Benotung)

Zu Semesterende können summative MC-Tests zur Beurteilung des erworbenen Wissens (Lernerfolg) eingesetzt werden. Hier geht es nicht mehr darum, Wissen zu vermitteln, sondern um eine möglichst objektive Messung des Wissensstandes und eine Gliederung der Ergebnisse in Form einer Rangfolge. Summative MC-Tests haben zwei Dimensionen: Sie sollen die Unterschiede zwischen den einzelnen Wissensständen ermitteln und den Erwerb von Kompetenzen messen. Hierarchisierung und Validität sind die Schlüsselbegriffe in diesem Zusammenhang.

Ein Test ist angemessen **hierarchisiert**, wenn sich anhand dieses Tests die Ergebnisse der einzelnen Studierenden nach Rängen gliedern lassen. Idealerweise verteilen sich die Ergebnisse in Form einer Gauss-Kurve (Normalverteilung). Ist der MC-Test insgesamt zu leicht zu beantworten, tendiert die Verteilungskurve nach rechts (Cap-Effekt), ist er zu schwer, tendiert sie nach links. Die Form der Kurve gibt an, ob ein Test die Ergebnisse erfolgreich differenziert. Sie sagt aber nichts über den Zusammenhang zwischen dem Schwierigkeitsgrad des Tests und der erreichten Lernziele aus (Validität).



Der oben erwähnte Cap-Effekt kann durchaus bedeuten, dass der betreffende Jahrgang von Studierenden homogener und kompetenter ist als andere Jahrgänge. Das Hierarchisierungskonzept führt zu einem moralischen Problem. Wenn Sie die Ergebnisse anhand einer Normalverteilung klassifizieren und jedem/jeder Studierenden einen Buchstaben oder eine Zahl zuweisen, bilden Sie jede einzelne Person nur in einer Dimension ab und geben unvollständige Informationen wieder. Als Lehrperson erteilen Sie Noten, die auf den Fähigkeiten (logisches Denken, Ausdrucksfähigkeit, ...) und dem (fachlichen oder sonstigen) Können der betreffenden Person beruhen und diese zusammenfassen. Umgekehrt ist es jedoch nicht möglich, von der Note auf die Fähigkeiten zu schliessen, auf denen sie beruht. Dies sollte man weder bei der Ausarbeitung von Tests noch bei der Interpretation der Ergebnisse vergessen.

Ein Test ist **valid**, wenn er effektiv das misst, was gemessen werden soll. Ein summativer MC-Test kann die Summe des erworbenen kognitiven, methodologischen und technischen Wissens der einzelnen Studierenden nur validieren, wenn die Fragen sich auf die im Semester erarbeiteten Lernziele beziehen. Er enthält weder Hinweise noch Mehrdeutigkeiten. Jeder nicht valide Test bestätigt die nachstehenden Aussagen von Kalz (in: Dessus und Tarbouriech, 2019): «Alle Studierenden, die mit ihrem Bildungswesen vertraut sind, wissen, dass sich schlechter Unterricht leicht vermeiden lässt, schlechte Beurteilungen dagegen kaum» (S. 91) [Übersetzung]. Ein Test kann valide sein und keine Standardabweichungen aufweisen, wenn sämtliche Studierenden den Stoff voll und ganz beherrschen und jede/r von ihnen ein ausgezeichnetes Ergebnis erzielt. Somit ist es wichtiger, dass Sie valide Tests durchführen, als dass diese Tests eine angemessene Hierarchisierung aufweisen.

Sie haben die Wahl zwischen verschiedenen **Fragetypen**: MC-Fragen, Wahr/Falsch-Fragen, Zuordnungsfragen (Verbindung von zwei gegebenen Informationen), Fragen mit numerischer Antwort und Kurzantworten-Frage. Sie können die Korrektur von Kurzantworten automatisieren, indem Sie alle Antwortmöglichkeiten vorgeben. Bei MC-Tests mit diagnostischem oder formativem Zweck bzw. tutorieller Begleitung empfehlen wir, nach Möglichkeit Fragen mit automatisierter Korrektur zu verwenden, um die Autonomie der Studierenden zu fördern. Bei summativen MC-Tests sollten Sie auf Wahr/Falsch-Fragen verzichten (die Wahrscheinlichkeit, dass die Studierenden einfach raten, ist im Vergleich zu Fragen mit mehreren Antworten erhöht). Wir empfehlen stattdessen, Kurzantworten-Fragen in den Test einzubauen (Korrektur durch die Lehrperson): Hier haben die Studierenden Gelegenheit, persönlich Antworten zu formulieren, zudem sinkt das Risiko, dass geraten oder abgeschrieben wird.

Summative MC-Tests gehen mit der Befürchtung einher, dass sie den Wissensstand der Studierenden nicht adäquat wiedergeben und dass es sich bei den Ergebnissen mindestens teilweise um **Zufallstreffer** handelt. Wenn Sie den Zufallseffekt bei summativen MC-Tests so gering wie möglich halten möchten, stehen Ihnen verschiedene Mittel zur Verfügung:

- In statistischer Hinsicht sinkt das Risiko von Zufallsantworten, je mehr Antwortvorschläge pro Frage vorgesehen sind. Wir empfehlen, jede Frage mit 4–5 Antwortmöglichkeiten zu versehen.
- Sie können verlangen, dass die Studierenden angeben, wie **sicher** sie sich bei den einzelnen Antworten fühlen, und diese Angaben in die Benotung einbeziehen. Die Tabelle unten (Prosperi, 2015) liefert ein Beispiel, wie Sie Punkte je nach Sicherheit und Richtigkeit der Antworten zählen können.

Tabelle 1

Punktvergabe anhand der Genauigkeit der Antwort und des gewählten Sicherheitsgrads

		Sicherheitsgrad					
		0	1	2	3	4	5
Antwort	Korrekt	+13	+16	+17	+18	+19	+20
	nicht Korrekt	+4	+3	+2	0	-6	-20

- Die Art, wie Sie die Punkte vergeben (Benotung), trägt ebenfalls dazu bei, Zufallsantworten zu vermeiden, wenn falsche Antworten zu Abzügen (Strafpunkten) führen. Wenn die Studierenden damit rechnen müssen, dass falsche Antworten sie Punkte kosten, nehmen sie sich Zeit zum Überlegen und vermeiden das mit Zufallsantworten verbundene Abzugsrisiko.

Eine zweite Befürchtung im Zusammenhang mit summativen MC-Tests bezieht sich auf das Risiko, dass Studierende **betrügen**. Dieses Problem lässt sich mit verschiedenen Funktionen (auf Moodle oder anderweitig) eingrenzen:

- So kann mit einem Zufallsgenerator eine bestimmte Anzahl Fragen aus einer Fragendatenbank gewählt werden. Dies führt dazu, dass nicht alle Studierenden dieselben Fragen beantworten. Hierbei müssen Sie beachten, dass alle Fragen gleich schwierig sein sollten, um die Vergleichbarkeit der Ergebnisse zu gewährleisten.
- Sie können die Fragen auch nach dem Zufallsprinzip anzeigen lassen, sodass die einzelnen Studierenden die Fragen in unterschiedlichen Anordnungen sehen.
- Die Anordnung der zur Auswahl stehenden Antworten lässt sich ebenfalls nach dem Zufallsprinzip anzeigen.
- Anders als bei MC-Tests mit formativem Zweck oder tutorieller Begleitung empfiehlt es sich hier, den einzelnen Studierenden die korrekte Antwort während der zeugnisrelevanten Beurteilung nicht anzuzeigen und auch kein paralleles Feedback anzubieten. Ein solches Feedback kann zu einem späteren Zeitpunkt erfolgen.
- Moodle bietet Ihnen die Möglichkeit, MC-Tests nur für eine bestimmte Zeit (eine Stunde oder eine halbe Stunde) zur Verfügung zu stellen, um die Kommunikation zwischen den Studierenden zu beschränken und Online-Suchstrategien zu verhindern.
- Setzen Sie Kurzantworten-Fragen ein, um persönliche Antworten der Studierenden zu erhalten.

Die zentrale Frage bei summativen MC-Tests ist die Frage der **Benotung**. Bei MC-Tests mit diagnostischem, formativem oder Bildungszweck spielt die Benotung kaum eine Rolle, da diese Fragen als Lerninstrumente eingesetzt werden. Geht es hingegen darum, die erworbenen Kompetenzen der einzelnen Studierenden zu überprüfen, sind Noten unentbehrlich.

Bei der Erstellung der Notenskala sollten Sie sich eine Reihe von Fragen stellen:

- Sind mehrere Antworten möglich oder nur eine?
- Werden alle Antworten auf dieselbe Weise gewertet oder haben gewisse Antworten einen höheren Stellenwert als andere? Gibt es Pass/Fail-Fragen? Die Gewichtung der Fragen hängt von der Gewichtung der Lernziele ab: Welches Konzept steht im Mittelpunkt? Welches ist schwieriger oder weniger wichtig? Sie haben die Wahl: Sie können entweder mit einer einzigen Notenskala für die gesamte Prüfung arbeiten oder nach Art der Fragen (längere, kompliziertere oder wichtigere Fragen ergeben mehr Punkte) bzw. nach Art der Aussagen gewichten (ein bestimmter schwerwiegender Fehler führt zu einem Abzug, ein anderer nicht).
- Führen falsche Antworten zu Abzügen? Wenn ja, wie gestalten sich die Abzüge?
- Führen fehlende Antworten zu Abzügen?

Die Tabelle unten (in Anlehnung an Bouvy und Vanhove, 2005) stellte die verschiedenen Benotungsansätze dar, massgeblich ist, ob eine oder mehrere korrekte Antworten möglich sind.



Klassische Notenskalen für MC-Tests ($k = \text{Anzahl Lösungsvorschläge}$, $c = \text{Anzahl korrekter Lösungen}$)

Skalen	Punktzahl				Vorzüge	Nachteile	
	Eine korrekte Antwort	Antwort korrekt		Keine Antwort			Nicht korrekte Antwort
		Alle korrekten Antworten	Ein Teil der korrekten Antworten				
Einfach	Eine korrekte Antwort	+1		0	0	Leicht zu handhaben Empfiehlt sich bei MC-Tests mit diagnostischem, formativem oder Bildungszweck Sie können für jede korrekte (Teil-)Antwort Punkte vergeben oder nur dann Punkte vergeben, wenn eine Frage vollständig beantwortet ist bzw. alle Antworten zu einer Frage vorliegen.	Wenig subtile Differenzierung der Kandidierenden. Begünstigt Antworten nach dem Zufallsprinzip. Keine Eignung für summative MC-Tests.
	Mehrere korrekte Antworten	+1	+1/c	0	0		
Symmetrisch	Eine korrekte Antwort		+1			0	-1
	Mehrere korrekte Antworten	+1	+1/c	0	-1/c		
Doppelt	Eine korrekte Antwort	+1		0	-0,5	Weniger ausgeprägte Sanktionierung von Zufallsantworten; ideal geeignet für summative MC-Tests	Nicht unbedingt populär
	Mehrere korrekte Antworten	+1	+1/c	0	-0,5/c		
Mit Einbezug von fehlenden Antworten	Eine korrekte Antwort	+1		+1/k	0	Weniger Zufallsantworten, da zufällig gewählte falsche Antworten sanktioniert werden. Ideal geeignet für summative MC-Tests.	Nicht unbedingt populär.
	Mehrere korrekte Antworten	+1	+1/c	+1/k	0		
Die Notenskala kann an Niveau/Anzahl der Antworten angepasst werden	Die Anzahl Punkte hängt von der Gesamtzahl der Fehler, nicht der Anzahl/Gesamtzahl der Fragen ab.		Kandidierende, die mehr Fehler machen, werden schlechter benotet als solche mit weniger Fehlern.		Die Fehler werden quantifiziert; ihr Inhalt spielt keine Rolle.		
Relative Konformität	Ergebnis = Anzahl korrekte Antw. d. einzelnen Kandidierenden / Gesamtzahl korrekte Antw. aller Kursteilnehmenden * Punktzahl pro Frage + Anzahl Fehler d. Kandidierenden / Gesamtzahl Fehler * Strafpunkt		Wenn die Punktzahl pro Frage der Anzahl Strafpunkte entspricht, ist die Notenskala ausgewogen.		In dieser Skala ist es nicht möglich, dass Kandidierende keine korrekten Antworten erzielen bzw. keine Fehler machen.		
Relative Konformität mit Antwortgewichtung	Ergebnisse = \sum der Koeffizienten der korrekten Antworten d. Kandidierenden / \sum der Koeffizienten der korrekten Antworten insgesamt * Punktzahl pro Frage + \sum der Koeffizienten der Fehler d. Kandidierenden / \sum der Koeffizienten der Fehler insgesamt * Strafpunkte		Antwortvorschläge lassen sich auf diesem Weg mehr oder weniger stark gewichten: ein gravierender Fehler oder eine grundlegende Antwort werden stark gewichtet und umgekehrt.				

3. Ausarbeitung

3.1 Die Anleitung

Die Anleitung beschreibt den Test, die Art der Bewertung, die Arten von Antworten bzw. die Erwartungen der Lehr- und Prüfungspersonen. Anleitungen werden nicht selten vernachlässigt - zu Unrecht, da sich die Prüfungen damit für die Studierenden intransparent und unangenehm gestalten. Studierende, die nicht wissen, was von ihnen erwartet wird und wie sie beurteilt werden, sind nicht in der Lage, ihre gesamten kognitiven Ressourcen zur Beantwortung der MC-Fragen einzusetzen. Das Hauptverdienst von Anleitungen liegt darin, dass sie Missverständnisse ausräumen und für mehr Transparenz sorgen:

- Geben Sie die Anzahl korrekter Antworten an («nur eine», «mehrere Antworten möglich», oder «nur eine, sofern in der Frage nichts anderes angegeben ist»)
- Erläutern Sie die Grundsätze der Benotung: Geben Sie die Notenskala für den MC-Test an (werden falsche bzw. fehlende Antworten mit Strafpunkten/Abzügen belegt, werden bestimmte Fragen anders gewichtet als der Rest, gibt es Pass/Fail-Fragen, müssen alle Antworten korrekt sein, wenn mehrere Antworten möglich sind, ... ?)
- Zweck des MC-Tests: Es empfiehlt sich, den Zweck des MC-Tests klarzustellen – geht es darum, den Kenntnisstand zu ermitteln (Diagnose), Gelerntes zu konsolidieren (Instruktion, tutorielle Begleitung), den Wissensstand zu ermitteln (formative Evaluation) oder die Studierenden zu benoten (summative Evaluation)?

Anleitungen sind vorzugsweise schriftlich niedergelegt, transparent und für den gesamten MC-Test gültig.

3.2 Die Problemstellung

Bevor Sie eine Problemstellung formulieren, sollten Sie sich vergegenwärtigen, welches Ziel Sie anstreben. Konzentrieren Sie sich auf ein einziges Ziel pro Problemstellung. Achten Sie auf Folgendes:

- Konzise und vollständige Problemstellungen, grundsätzlich frei von überflüssigen Angaben;
- Die für die Antwort erforderlichen Angaben sind in die Problemstellung aufzunehmen, nicht in die Antwortauswahl;
- Die Problemstellung sollte keinen Anlass für Wahr/Falsch-Antworten geben;
- Die Problemstellung sollte nach Möglichkeit keine Verneinungen enthalten. Die Frageform ist möglich: Sie wirkt anregend auf die Studierenden und vereinfacht die Beantwortung (weniger grammatikalische Bedenken, weniger Hinweise, dass eine bestimmte Antwort korrekt ist). Mit Aussagesätzen lässt sich eine gewisse Abwechslung erzielen, sodass nicht alle Fragen dieselbe Struktur haben.

Vollständige Problemstellung – Beispiel:

Im Jahr 1898 richtet Émile Zola einen offenen Brief an den französischen Präsidenten. Seine Anklage «J'accuse...!» richtet sich gegen die Verurteilung des Hauptmanns Alfred Dreyfus. Welche Zeitung verwendete «J'accuse ...!» als Schlagzeile?

a. Le Matin

b. Aujourd'hui

c. L'Aurore

3.3 Die korrekte(n) Antwort(en) (bzw. Antwortmöglichkeiten)

- Die korrekten Antworten müssen zweifelsfrei **zutreffen**.
- Sie werden **den Distraktoren** nach Möglichkeit **angepasst** (Länge, Aufbau, Präzision, ...), da formale Unterschiede ansonsten als Hinweis auf die richtige Antwort dienen könnten.
- Die Reihenfolge der Distraktoren und korrekten Antworten pro Frage sollte im gesamten Test nach dem Zufallsprinzip abwechseln. Bei bestimmten Fragen können Sie eine interne logische Reihenfolge einhalten (z. B. alphabetische, numerische oder chronologische Reihenfolge).
- Sie können das Problem so stellen, dass die korrekte Antwort «keine dieser Antworten» lautet. Auf die Antwortmöglichkeit «alle diese Antworten» sollten Sie verzichten, wenn die Studierenden wissen, dass es nur eine korrekte Antwort gibt. Studierende, die zwei korrekte Antwortmöglichkeiten kennen, wählen in diesem Fall die Option «alle diese Antworten».
- Vermeiden Sie konvergierende Antworten (die Antwort, welche die meisten Elemente mit den anderen gemeinsam hat, dürfte mit hoher Wahrscheinlichkeit korrekt sein) oder zwischen der Problemstellung und der korrekten Antwort (die Antwort, welche die meisten Schlüsselbegriffe aus der Problemstellung aufnimmt, weckt allenfalls den Eindruck, sie sei korrekt). Vermeiden Sie nach Möglichkeit Hinweise, welche 2-3 Antwortmöglichkeiten ins Zentrum der Aufmerksamkeit rücken (wenn sich zwei Antwortmöglichkeiten sehr ähnlich sehen, besteht die Gefahr, dass die anderen in den Hintergrund gedrängt werden).

Beispiel für eine korrekte Antwort, die den Distraktoren ähnlich sieht

Welcher Begriff bezeichnet den Erlös eines Verkäufers, der über den kostendeckenden Betrag hinausgeht?

- a) Die Konsumentenrente
- b) Die Produzentenrente
- c) Der Nettogewinn
- d) Der Nettobetriebsüberschuss

3.4 Die Distraktoren (Fallen/Fangfragen, falsche Antwortmöglichkeiten)

Die falschen Antwortmöglichkeiten werden als Distraktoren bezeichnet. Ein Grossteil der Schwierigkeiten beim Erarbeiten von MC-Tests lässt sich auf die Distraktoren zurückführen. Sie müssen:

- eindeutig falsch ... aber subtil angelegt sein. Im Idealfall beantwortet ein Distraktor nur einen Teil der Problemstellung oder er wirkt korrekt, bis man ihn genauer analysiert.
- Vermeiden Sie Distraktoren, die sich nur geringfügig voneinander unterscheiden.
- Geben Sie eine Reihe von Distraktoren an (mindestens 3), um das Risiko von Zufallsantworten möglichst gering zu halten.
- Gestalten Sie die Distraktoren so, dass sie der korrekten Antwort ähnlich sehen (Länge, Komplexität, Stil, Formulierung, ...)
- Vermeiden Sie Komplexitäten, die nicht mit der Fragestellung zusammenhängen.
- Verzichten Sie auf vage Begriffe (im Allgemeinen, manchmal, ...) und ebenso auf absolute Begriffe (immer, nie, ...)
- Sie können die Antwortmöglichkeit «Ich weiss nicht» vorsehen und die Sanktionen weniger streng gestalten, wenn diese Antwort gewählt wird. Für summative MC-Tests ist dieses Verfahren weniger geeignet. Das Eingeständnis der eigenen Unwissenheit ist häufig der erste Schritt zum Wissen.

Welche Rezeptoren ermöglichen die Wahrnehmung von Wärme?

- a. Krause-Endkolben
- b. Meissner-Körperchen
- c. Freie Nervenendigung
- d. Ruffini-Körperchen
- e. Golgi-Mazzoni-Körperchen
- f. Merkel-Scheibe
- g. Vater-Pacini-Körperchen

Aufmerksame Studierende entnehmen diesem Beispiel, dass sie ihre Lernarbeit auch auf die übrigen Hautrezeptoren (Druck, Schmerz, ...) ausweiten müssen. Johnstone (2003) geht davon aus, dass ein Distraktor dann sinnvoll ist, wenn rund 5% der Studierenden ihn als richtige Antwort missverstehen. Wird ein Distraktor von niemandem angekreuzt, können Sie ihn streichen. Vorlagen für Distraktoren finden Sie in den häufig wiederkehrenden fehlerhaften Antworten der Studierenden, in häufigen Verwechslungen und Irrtümern oder im kreativen Umgang mit gängigen unbestätigten Vorurteilen.

3.5 Formative Rückmeldung (Feedback)

Mithilfe der Plattform Moodle können Sie Feedbacks automatisieren. Diese Möglichkeit ist von grundlegender Bedeutung: So lässt sich ein formativer MC-Test in einen tutoriell begleiteten MC-Test umwandeln. Feedbacks zeigen den Studierenden auf, wo sie gescheitert sind und/oder was sie missverstanden haben. Sie beschleunigen den Lernprozess, indem sie die individuellen Schwierigkeiten benennen und verdeutlichen. Wenn die Studierenden das Ausmass ihres Unwissens nicht kennen, können sie auch keine Massnahmen ergreifen, um ihm abzuhelpfen. Feedbacks bestehen aus zwei Komponenten: Der Validierung der Antwort («korrekt» bzw. «nicht korrekt») und einer Erläuterung. Die Erläuterung spielt für uns eine wesentlichere Rolle: Sie zielt darauf ab, dass die Studierenden ihre Fehler nachvollziehen können.

Legen Sie formative erläuternde Feedbacks nutzbringend, kurz und vollständig an, d. h. passen Sie sie dem Niveau an, das Sie im Unterricht erwarten. Geben Sie häufig und **umgehend**¹ Feedback. Als Lehrperson haben Sie folgende Möglichkeiten:

- Analog zum Vorgehen in Ihrem Unterricht können Sie Erklärungen geben. Sie können den zugehörigen Unterrichtsstoff einkopieren.
- Sie können ein Argument verdeutlichen oder Lücken in der Argumentation aufzeigen.
- Sie können schädliche Distraktoren ausschliessen und diesen Ausschluss begründen. Im Rahmen eines formativen MC-Tests lohnt es gegebenenfalls, ausnahmsweise nur verwirrende Distraktoren zu wählen und anschliessend die Verwirrung mit einem klärenden Feedback zu beheben. In diesem Fall dient der komplizierte Distraktor dazu, bei den Studierenden Neugierde oder Interesse zu wecken bzw. sie zu einer kritischen Auseinandersetzung mit dem Thema zu motivieren.

¹ In Moodle haben Sie die Wahl zwischen direktem Feedback nach jeder Frage oder Feedback nach Abschluss des gesamten Tests. Das direkte Feedback nach jeder Frage ist wirksamer, da sich so die falschen Vorstellungen der einzelnen Studierenden umgehend korrigieren lassen.

Rückmeldungen – zwei Beispiele

Im Gehör erfolgt die Signalübertragung über

- a. Fotorezeptoren
- b. Papillen
- c. Freie Nervenendigung
- d. Haarzellen
- e. Krause-Endkolben

Die korrekte Antwort lautet «Haarzellen»

oder

Antwort d. ist korrekt.

Langfristig handelt es sich bei allen Kosten um variable Kosten; Fixkosten gibt es langfristig nicht.

- a. Wahr
- b. Falsch

Die kurze und die lange Frist unterscheiden sich (in wirtschaftswissenschaftlicher Hinsicht) wie folgt voneinander: Kurzfristig sind bestimmte Produktionsfaktoren fix, während sich langfristig die Menge aller Produktionsfaktoren anpassen lässt. Dies bedeutet mit anderen Worten, dass langfristig alle Produktionsfaktoren variabel sind. Je nach Art des zu analysierenden Problems ist der Begriff Kurz- bzw. Langfristigkeit angebracht. Langfristigkeit bezeichnet den Zeitraum, den ein Unternehmen zur Anpassung derjenigen Produktionsfaktoren benötigt, die kurzfristig fix sind. Bei den fixen Kosten des Unternehmens handelt es sich um die Entgelte für Faktoren, welche das Unternehmen kurzfristig nicht anpassen kann und die somit nicht vom Produktionsniveau abhängen.

Im ersten Beispiel schafft das Feedback keinen Mehrwert für die Studierenden, während es sich beim zweiten Feedback um eine lehrreiche Rückmeldung handelt. Dies soll nicht heissen, dass das erste Beispiel schlecht ist: Die Lehrperson präzisiert nicht systematisch, sondern wählt gezielt Fragen aus, bei denen sie der Ansicht ist, dass die entsprechenden Rückmeldungen nützlich und wertvoll sind.

Schlechte Praxis: Unnützes Feedback, ein Beitrag von Fernand Raynaud, 1976:

Toto: Ich hab da eine Frage: Wie funktionieren Atom-U-Boote überhaupt unter Wasser? Weil ... einerseits haben sie keinen Dieselmotor, auch keinen Elektromotor ... In der Pause haben wir es gerade noch davon gehabt ... Wieso funktionieren sie überhaupt unter Wasser, Papi?
Vater: Ähhhh ... hmhhh... mein lieber Junge... sie sind dazu gemacht!

Wie viele gefühlte Scheinerklärungen, wie viele gefühlte «Ähhhh ... hmhhh ... mein Lieber ... sie sind dazu gemacht! », haben die Studierenden wohl während ihrer Bildungslaufbahn gehört?

Fazit

Nützlich Feedback ermöglicht den Lernenden, die Lücke zwischen ihrem Wissensstand und dem von der Lehrperson erwarteten Wissensstand zu erkennen. Die als Rückmeldung gelieferten Informationen tragen dazu bei, dass sie diese Lücke schliessen können.

4. Vorgefasste Meinungen

Im Folgenden befassen wir uns mit vorgefassten Meinungen zum Thema MC-Tests. Zumeist handelt es sich um Befürchtungen, die Sie vielleicht ebenfalls teilen. Zunächst möchten wir aufzeigen, inwiefern diese Meinungen und Vorurteile wahr bzw. falsch oder gar beides sind.

4.1 Vorgefasste Meinung Nr. 1: Treffer können zufallsbedingt sein

WAHR: Beim Korrigieren lässt sich nicht zwischen «zufällig korrekten» und «auf korrektem Wissen beruhenden» Antworten differenzieren. Dieses Problem verliert mit steigender Anzahl an Distraktoren an Bedeutung (lässt sich aber nicht vollständig ausschliessen). Sie können es ausschliessen, wenn Sie ein anderes Format als MC-Tests verwenden und die Lernenden schriftliche Antworten ausarbeiten lassen.

FALSCH: Ein einziges Beispiel genügt, um diese vorgefasste Meinung zu widerlegen (René Krebs, 2004):

Ein 47-jähriger Patient sucht die Notfallstation auf, weil er mit dem Velo gestürzt ist. Er wirkt aufgeregt und weist einen intensiven Foetor alcoholicus auf. Er klagt über thorakale Schmerzen im linken unteren Bereich und im linken Hypochondrium. Sein Wert auf der Glasgow-Skala liegt bei 15. Es sind keine äusserlichen Wunden zu sehen. Die Atemfrequenz beträgt 38, die Pulsfrequenz 140, der Blutdruck liegt bei 90/60 mmHg. Wie lässt sich das hämodynamische Bild am glaubhaftesten erklären?

- (A) durch einen kardiogenen Schock
- (B) durch eine Hypovolämie
- (C) durch eine Lungenembolie
- (D) durch eine Folge der Alkoholvergiftung
- (E) durch einen posttraumatischen vagalen Reflex

Hier zeigt sich, dass die sorgfältige Auswahl der Problemstellung und der Distraktoren dazu führt, dass Studierende, die den Stoff nicht ausreichend beherrschen, die Frage nicht lösen können. Angemerkt sei ferner, dass die Beantwortung dieser Frage eine hoch entwickelte kognitive Aufnahmefähigkeit voraussetzt: Die Studierenden müssen die Angaben synthetisch erfassen und ihre Kenntnisse anwenden.

WAHR: Wenn Sie als Lehrperson zwei Antwortmöglichkeiten pro Frage vorsehen und pro korrekter Antwort einen Punkt sowie null Punkte für fehlerhafte Antworten vergeben, decken sich die Ergebnisse nicht mit den tatsächlichen Kenntnissen der Studierenden. Dann verhält es sich nämlich so, dass Studierende, die systematisch nach dem Zufallsprinzip vorgehen, die Hälfte der Punkte erhalten (d. h. die Anzahl Punkte, die der Note 3,5 entsprechen). Zudem tendiert die Verteilungskurve asymmetrisch nach rechts (Cap-Effekt). Sie können die Zufallseffekte einschränken, indem Sie die Anzahl Antwortmöglichkeiten auf 3 und anschliessend auf 5 erhöhen, trotzdem wird ein Rest der Note auch weiterhin zufallsbedingt sein.

Diese Art der Gewichtung bei strikter Übereinstimmung ist leicht anwendbar und demzufolge beliebt. Es handelt sich um eine unanfechtbare Notenskala für summative Bewertungen.

FALSCH: Der oben dargestellte Mangel lässt sich mit verschiedenen **Benotungsansätzen** beheben, die Zufallseffekte vereiteln sollen ([siehe Seite 8](#)). Diese Methoden beziehen die Anzahl Distraktoren ein und stützen sich auf mehr oder weniger komplizierte Berechnungen, daher sei an dieser Stelle eine einfache Alternative vorgeschlagen.

Bevor wir die Methode an sich vorstellen, möchten wir die zwei Funktionen eines summativen Tests noch einmal in Erinnerung rufen: Ein solcher Test soll das messen, was zu messen ist (Validität), und die Ergebnisse hierarchisieren. Ein Ansatz besteht in der Definition einer Untergrenze, von der aus die Bruttopunkteergebnisse nach vorab definierten Einteilungen **hierarchisiert** werden. Zwischen der Untergrenze (0 Punkte) und dem höchstmöglichen erreichbaren Wert werden die Punktzahlen linear verteilt.

Nehmen wir an, dass ein Test 20 Fragen umfasst. Sie können nun einen Mindesterfolg bei 10 korrekten Antworten festlegen. Dann legen Sie linear folgende Werte fest:

- | | |
|--|-----------|
| - 20 korrekte Antworten: Höchstwert | 10 Punkte |
| - 18 korrekte Antworten: | 8 Punkte |
| - 15 korrekte Antworten: in der Mitte zwischen Mindest- und Höchstwert | 5 Punkte |
| - 10 korrekte Antworten: Mindestwert | 0 Punkte |
| - 5 korrekte Antworten: ungenügend | 0 Punkte |

Mit diesem Verfahren lassen sich Benotungen von Zufallstreffern teilweise vermeiden; es führt zu einer Differenzierung der Ergebnisse der einzelnen Studierenden, es lässt sich leicht einrichten und überlässt die Festlegung der Grenzen der Lehrperson. Es dürfte aber kaum beliebt sein, da Studierende, die nur 1 von zehn Punkten erhalten, obwohl sie 11 Fragen richtig beantwortet hatten, mit ziemlicher Sicherheit enttäuscht sind. Diese Studierenden verdrängen die Tatsache, dass ihr Ergebnis die Folge davon ist, dass sie den Stoff nur teilweise beherrschen. Wahrscheinlich waren sie sich bei einigen Antworten ihrer Sache sicher, bei anderen sind sie nach dem Ausschlussverfahren vorgegangen und bei weiteren Fragen haben sie auf den Zufall vertraut.

Als Lehrperson tragen Sie die Verantwortung für die Wahl Ihres Benotungssystems. Dabei sollten Sie nicht vergessen, dass das Risiko einer Verzerrung der Ergebnisse beträchtlich ist.

4.2 Vorgefasste Meinung Nr. 2: Die korrekte Antwort lässt sich durch Ausschluss der Distraktoren ableiten.

WAHR: Wie viele Studierende haben Noten erzielt, die ihrem Wissensstand nicht entsprachen, indem sie bei manchen Fragen die Antworten einfach geraten haben? Besser gesagt, bei wem war das nicht der Fall? Wie viele Studierende sind auf Distraktoren oder eine komplizierte Formulierung hereingefallen, obwohl sie die korrekte Antwort wussten? Besser gesagt, bei wem war das nicht der Fall?

FALSCH: Wenn die Distraktoren sorgfältig ausgewählt wurden, lässt sich die Antwort nicht ableiten.

<u>Negativbeispiel</u>	<u>Positivbeispiel</u>
<p>Der Blauwal gehört zu den längsten lebenden Tieren der Welt. Wie lang ist er wohl?</p> <ul style="list-style-type: none"> - Über 3 Kilometer - Über 30 Meter lang - Über 30 Zentimeter lang - Über 300 Meter lang <p><i>Nur wer die Längenmasse nicht kennt, kann die korrekte Antwort nicht ableiten.</i></p>	<p>Der Blauwal stellt Grössen- und Gewichtsrekorde auf. Welche der nachfolgenden Aussagen ist korrekt?</p> <ul style="list-style-type: none"> - Über 30 Meter, aber unter 170 Tonnen. - Über 40 Meter und über 170 Tonnen. - Über 40 Meter, aber unter 170 Tonnen. - Über 30 Meter und über 170 Tonnen. <p>Die Distraktoren sorgen dafür, dass sich die korrekte Antwort nicht einfach ableiten lässt.</p>

FALSCH: Wenn ein Lösungsweg für die Antwort erforderlich ist (Berechnung, Überlegung), können die Studierenden sie nicht im Ausschlussverfahren ermitteln. Beispiel: «Ein Mann ist viermal so alt wie sein Sohn. In zwanzig Jahren wird er doppelt so alt sein wie sein Sohn. Wie alt ist er heute? »

Leider FALSCH: Jedes Mal, wenn Sie als Lehrperson eine Fangfrage stellen und die Studierenden dies nicht merken. Die Studierenden antworten automatisch falsch, weil ihre kognitive Leistung nicht tiefgreifend genug ist. Solche Automatismen sind ein häufiger Grund für Schnitzer. Die Studierenden behandeln die Frage cursorisch und beantworten die Frage, die sie anscheinend verstanden haben, anstatt sich mit der Bedeutung der Frage auseinanderzusetzen. Mit den Worten von Borst G. (2016):

Es ist von grundlegender Bedeutung, einem Kind [einem/einer Studierenden] Fangfragen/Fallen aufzuzeigen, damit es diese vermeiden kann. Mithilfe der funktionellen Bildgebung haben wir belegt, dass das Gehirn von Fehlern zu Erfolgen übergeht, indem es sich neu konfiguriert. Die Aktivierung des Gehirns verlagert sich aus dem hinteren – an Automatismen beteiligten – Bereich der Hirnrinde in den präfrontalen Cortex, d. h. den Bereich, der die Heuristik blockiert.

Die Studierenden sollten imstande sein, ihre automatischen Antwortmechanismen zu blockieren, um ihre in die Tiefe zu denken. Der Hinweis «Achtung, Fangfrage» reicht aus, um Studierende aus automatischen Gedankengängen aufzurütteln und sie zu eigentlichen Überlegungen zu bewegen und damit sie aus einer passiven Haltung zu einem aktiven Lernen finden.

4.3 Vorgefasste Meinung Nr. 3: GMC-Tests zu entwerfen ist zeitaufwändig

WAHR: Sie benötigen Zeit, um einen MC-Test zu entwerfen und zu erstellen. Dies gilt insbesondere für summative MC-Tests. Bei formativen oder tutoriell begleiteten MC-Tests hat diese Aussage nur beschränkt Gültigkeit, dies aus vier Gründen:

- Ein formativer MC-Test muss nicht derart perfekt durchgestaltet sein wie ein summativer MC-Test, von dem schliesslich die Noten der Studierenden abhängen.
- Moodle bietet Ihnen die Möglichkeit, Feedbacks und Hilfestellungen zu erfassen, d. h. nicht perfekt formulierte Fragen zu nuancieren. Ungeschickte Formulierungen oder mehrdeutige Distraktoren lassen sich mit einem klärenden Feedback oder einem Hinweis als Hilfestellung präzisieren.
- Formative oder tutoriell begleitete MC-Tests können von Jahr zu Jahr erneut eingesetzt (und somit auch verbessert oder angereichert) werden.
- Wenn Sie als Lehrperson in der Anleitung die Studierenden auffordern, ebenfalls zur Verbesserung der MC-Tests beizutragen (indem sie auf Fehler und Tippfehler aufmerksam machen und Anmerkungen anbringen), erhalten Sie einerseits nützliche Rückmeldungen und andererseits halten Sie Kontakt mit Studierenden, die Sie nur selten physisch treffen.

FALSCH: Wenn Sie den Zeitstrahl bis zur Korrekturphase weiterziehen, sehen Sie, dass Sie durch automatisierte Feedbacks und das selbstgesteuerte Lernen der Studierenden Zeit gewinnen. Diese Zeit steht Ihnen als Lehrperson für andere Zwecke zur Verfügung.

FALSCH: Einmal formuliert, lassen sich zahlreiche Fragen in (leicht) abgewandelter Form wiederverwenden

4.4 Vorgefasste Meinung Nr. 4: Mit MC-Tests lässt sich nicht messen, ob die Studierenden in der Lage sind, analytisch zu denken oder vom Stoff zu abstrahieren.

IRREFÜHREND Diese vorgefasste Meinung hält sich hartnäckig, mit heimtückischer Wirkung. Die Frage ist legitim, setzt aber am falschen Ort an. Es geht nicht darum, ob die Studierenden Konzepte kennen, verstehen und anwenden oder Urteile fällen müssen. Es geht auch nicht darum, ob der Wissenserwerb oder aber die Urteilsfähigkeit im Zentrum stehen soll. Es geht darum zu wissen, wie Wissen erworben werden kann UND wie Urteile gefällt werden können, «denn ohne Wissen verfügt der Mensch nicht über die notwendigen Grundlagen für die Entwicklung von übergeordneten Analyse-, Synthese- und kreativen Denkfähigkeiten» (Brown P.C. et al., S. 39).

SIE HABEN DIE WAHL: Dieses Vorurteil lässt sich mit zwei Ansätzen entkräften, die zu vergleichbaren Schlussfolgerungen führen: einem kognitiven Ansatz (1) und einem pädagogischen Ansatz (2):

1. Bei der Entwicklung ihrer Fertigkeiten setzen die Studierenden zahlreiche mentale Prozesse und kognitive Fähigkeiten (Gedächtnis, Logik, Problemlösung, Entscheidungsfindung, Aufmerksamkeit, ...) ein.
2. In der Pädagogik wird die Klassifizierung üblicherweise anhand von [Niveaustufen der kognitiven Prozessebenen](#) des Wissenserwerbs vorgenommen: Wissen, Verstehen, Anwenden (Analyse), Synthese und Bewertung (Evaluation). (Taxonomie von Bloom).

Im Zusammenhang mit der Konzeption von MC-Tests stellen sich somit zwei Fragen:

1. Beschränken sich MC-Tests auf eine teilweise Mobilisation der kognitiven Fähigkeiten und führen sie deshalb zu einem fragmentarischen oder oberflächlichen Wissenserwerb?
2. Ermöglichen MC-Tests einen Wissenserwerb auf höheren Niveaustufen (insbesondere auf den Ebenen Analyse, Evaluation und Synthese)?

In beiden Fällen ist Ihnen als Lehrperson daran gelegen, den oberflächlichen Wissenserwerb ohne Tiefgang zu vermeiden. Als Lehrperson ist Ihnen mehr daran gelegen, dass Bürgerinnen und Bürger in der Lage sind zu nuancieren, als daran, Gelehrte heranzuziehen. Gewissenhafte Lehrpersonen folgen den Spuren von Montaigne:

Der Lehrer sollte von seinem Schüler nicht nur verlangen, dass er die Lektion im Wortlaut wiedergibt, sondern dass er die Bedeutung und die wesentlichen Inhalte wiedergeben kann. Den Nutzen der Lektion sollte er nicht anhand der Gedächtnisleistung beurteilen, sondern anhand des Verhaltens [des Schülers]. [...] Wer sein Essen im gleichen Zustand wieder auswürgt, in dem er es hintergeschluckt hat, zeigt, dass es nicht verdaut wurde: Der Magen hat seine Arbeit nicht getan, wenn er den Zustand und die Form der ihm zur Verdauung zugeführten Nahrung nicht verändert. (Essais I,26).

Sind MC-Tests geeignet, den Unterricht mit Sinn und Inhalt anzureichern? Tragen MC-Tests dazu bei, dass sich der Zustand und die Form des Lernstoffs ändern? Lassen sich mit MC-Tests die analytischen und die Anwendungsfähigkeiten der Studierenden messen? Wie gerne würden manche diese Fragen dreimal mit Nein beantworten – trotzdem sollte man sich vor oberflächlichen Schlussfolgerungen hüten. Nicht alle MC-Tests messen den Wissensstand nur unvollständig, indem sie sich einseitig auf Gedächtnisleistungen ausrichten, denn sie können auch andere kognitive Niveaus mobilisieren. Im Folgenden nehmen wir drei Beispiele genauer unter die Lupe:

	Problemstellung	Ordnen Sie die vorgeschlagenen Definitionen den Optionen zu
Frage 1	Ordnen Sie jeder Lapsus-Definition (Fromkin, 1973) die entsprechende Bezeichnung zu.	a) Der/die Sprechende ersetzt einen Begriff durch einen anderen mit ähnlichem Klang, aber unterschiedlicher Bedeutung. b) Der/die Sprechende ersetzt einen sprachlichen Bestandteil durch einen anderen. c) Der/die Sprechende kehrt die Stellung von zwei sprachlichen Bestandteilen um. d) Der/die Sprechende verwendet einen sprachlichen Bestandteil in einem anderen als dem korrekten Satzteil, da dieser Bestandteil in der Aussage erst später verwendet werden sollte. e) Der/die Sprechende verwendet einen sprachlichen Bestandteil, der im früheren Satzverlauf angebracht gewesen wäre, anschliessend jedoch nicht mehr passt.	<ul style="list-style-type: none"> • Perversion • Inversion • Substitution • Antizipation • Inkongruenz.
	Diese Problemstellung beansprucht die Fähigkeit zu logischer Überlegung, die Aufmerksamkeit und das Gedächtnis ... Sie ist auf dem kognitiven Niveau der Anwendung angesiedelt: Da die Lernenden anhand einer Regel Schlussfolgerungen ziehen müssen, geht die gefragte Leistung über das Wissens- und das Verstehensniveau hinaus. So verhält es sich bei jeder Aufgabe, in der ein Beispiel auf eine Regel zurückzuführen ist.		

	Problemstellung	Antwortmöglichkeiten
Frage 2	In der Linguistik wird zwischen <i>Lexik</i> und <i>Vokabular</i> differenziert. Der Grund dafür ist folgender:	a) Beide Begriffe können als Synonyme betrachtet werden. b) Die Lexik bezeichnet die Gesamtheit der Morpheme einer gegebenen Sprache, während das Vokabular die im Repertoire vorhandenen Wörter bezeichnet. c) Der Begriff Lexik bezeichnet die Gesamtheit der von einer Person verwendeten Wörter, während der Begriff Vokabular die Gesamtheit der Wörter einer Sprache bezeichnet.
	Diese Frage bezieht sich auf einen bestimmten Kontext (Linguistik). Die Lernenden werden somit aufgefordert, ein bestimmtes Wissen im Kontext einer gegebenen Situation anzuwenden . Wenn Sie als Lehrperson eine Frage mit «Der/die Wissenschaftler/in XYZ (Jahr) ist der Ansicht, dass ...» einleiten, verlangen Sie von Ihren Lernenden mehr als eine blosser Anwendung: Hier ist die Fähigkeit zur Analyse bzw. Beurteilung gefordert.	

	Problemstellung	Ordnen Sie die vorgeschlagenen Definitionen den Optionen zu:
Frage 3	Bitte ordnen Sie jeder Schwachstelle die entsprechende Forschungsmethode aus der Liste zu:	a) Unmöglichkeit, unbewusste Prozesse zu beschreiben. Die Selbstbeobachtung der kognitiven Tätigkeit kann ihrerseits zu einer Verzerrung der Beobachtung führen. b) Mangelnde Nachprüfung durch Experimente, Verzerrung der Beobachtungen. c) Es ist nicht immer möglich, die Ergebnisse über einen bestimmten Ort, eine bestimmte Zeitdauer und eine spezifische Aufgabe hinaus zu verallgemeinern; Abweichungen zwischen dem Verhalten im Alltag und demjenigen im Labor. d) Beschränkungen aufgrund der Hardware-Begrenzungen, Problem bei der Verallgemeinerung der Ergebnisse auf das natürliche Erkenntnisvermögen e) Das Ergebnis lässt sich nicht verallgemeinern. f) Beschränkter Zugang für Forschende, Ausrüstung schwer zu finden und sehr kostspielig, Problem der Verallgemeinerung bei Beobachtung von Gehirnen, die von der Norm abweichen.	<ul style="list-style-type: none"> • Introspektive Berichte/Protokolle • Naturforschungsbeobachtungen • Fallstudien • Computersimulationen • Psychobiologische Forschung • Kontrollierte Laborexperimente
	<p>Was ist schwieriger, die separate Aufzählung der Schwachstellen der einzelnen Forschungsmethoden oder die Wahl zwischen den ungeordneten Antwortmöglichkeiten auf diesen beiden Listen? Die Antwort hängt von den Studierenden ab.</p> <p>Wir befinden uns hier auf dem Niveau der Anwendung / Analyse bzw. Synthese oder Evaluation (abhängig von der Einstellung der Antwortenden).</p> <p>Diese Fragen beanspruchen die Fähigkeiten Aufmerksamkeit, logische Überlegung und Entscheidungsfindung.</p>		

Wenn die Fragen dazu führen, dass das Wissen eingehender unter die Lupe genommen werden muss, beanspruchen MC-Tests durchaus Fähigkeiten auf hohem kognitivem Niveau. Sie können beispielsweise:

- Fragen zur Beziehung (zeitlich, logisch, ...) zwischen verschiedenen Konzepten stellen.
- Die Fähigkeit zur Anwendung von Kenntnissen in Situationen testen, die im Unterricht nicht behandelt wurden.
- Analogien aufführen, sodass die Studierenden gezwungen sind, Konzepte zu vergleichen und zwischen ihnen zu unterscheiden.
- Fragen stellen, bei denen verschiedene logische Schlussfolgerungen gezogen werden müssen, um die korrekte Antwort zu finden.
- Verschiedene kognitive Fähigkeiten einbinden, indem Sie «eine Evaluation des erworbenen Wissens verlangen, um zu einer logischen Entscheidung zu kommen» (Dessus und Tarbouriech, 2019).

Zusammenfassend lässt sich sagen, dass Sie als Lehrperson gut daran tun, Ihre Ansätze bei der Befragung der Studierenden zu diversifizieren und auf verschiedenen Komplexitätsebenen zu arbeiten. So können Sie beispielsweise komplexe Darstellungen auf Einzelfälle anwenden lassen oder die relative Rangfolge bestimmter Ideen bestimmen lassen.

5. Zusammenfassung: Fallstricke und Hinweise

Man kann das, was man weiss, nicht überlegt anwenden, wenn es nichts anzuwenden gibt. Sternberg R. et al. (2008).

- **Mangelnde Klarheit:** Klare, unzweideutige Formulierungen (lassen Sie Ihren MC-Testentwurf von einer anderen Person gegenlesen).
- **Inkohärente Granularität:** Wenn Sie Allgemeinwissen testen wollen, setzen Sie auf allgemeine Fragen, Detailfragen eignen sich zur Überprüfung von tiefgreifendem Wissen.
- **Eintönigkeit:** Sorgen Sie für Abwechslung, indem Sie die Fragen unterschiedlich gestalten (Länge, Thema, Kognitionsebene, ...), um die Anwendung von geschlossenen Lernstrategien (wie Auswendiglernen anstatt zu lernen) zu verhindern. Zögern Sie nicht, verschiedene Themenkreise anzusprechen, sofern sie untereinander eine Verbindung aufweisen. Es verhält sich effektiv so, dass sich Lernstoff leichter aneignen lässt, wenn er in beliebiger Reihenfolge angegangen wird, anstatt ein Thema nach dem anderen abzuhandeln (Brown P.C. et al., 2014).
- **Verzicht auf Anleitungen:** Erläutern Sie die Notenskala, die Zwecke, die Erwartungen und die Ziele.
- **Beschränkung auf die Niveaustufe «Wissen»:** Sie können die Studierenden dazu motivieren, analytisch oder synthetisch vorzugehen bzw. Urteile zu fällen, indem Sie Fragen stellen, die sich auf wissenschaftliche Kontroversen beziehen, Nuancen betreffen, unterschiedliche Standpunkte verschiedener Autoren anführen, Beispiele verwenden, die Theorien zuzuordnen sind, Zitate angeben, bei denen nach den Autoren gesucht wird, ...
- **Bekräftigung von falschen Ansätzen:** Die Studierenden widmen der Frage nur dann ihre gesamte Aufmerksamkeit, wenn sie sich ihres Nichtwissens bewusst sind. Ansonsten besteht die Gefahr, dass sie die Frage kursorisch durchlesen und rasch antworten. Wenn Sie für umgehendes Feedback sorgen, tragen Sie dazu bei, dass die Studierenden sich ihrer Wissenslücken bewusst werden. Es lohnt sich zudem, auf Fangfragen hinzuweisen: So nutzen die Studierenden bei der Antwort ihre gesamten kognitiven Fähigkeiten.
- **Übermässiger Aufwand bei der Parametrisierung in Moodle:** Es empfiehlt sich, die Fragen separat zu konzipieren und sie anschliessend in Moodle einzufügen. Ein Sonderfall wurde bis jetzt noch nicht behandelt: Es ist möglich, die Definition eines Konzepts in einem Moodle-Quiz zu erfragen. Die Studierenden definieren Ihren Leitgedanken, anschliessend wird ihnen Ihre Definition angezeigt. Die selbstständige Formulierung der im Unterricht behandelten Leitgedanken fördert den Wissenserwerb!
- **Zu enge Verwandtschaft der Fragen untereinander:** Wenn die Überschrift einer Frage die Lösung für eine weitere Frage birgt, ist der Test nicht länger valide. Vermeiden Sie also übergreifende Lösungshinweise und sorgen Sie in Moodle dafür, dass zwei allzu eng verwandte Fragen räumlich auseinander gesetzt werden.
- **Amüsement um jeden Preis:** Zwingen Sie sich nicht zu äusserst fantasievollen Formulierungen (sofern Sie nicht einen Heiterkeitseffekt beabsichtigen). Langeweile lässt sich am wirkungsvollsten durch eine geschickte Themenwahl, Fragen mit hohem kognitivem Anspruch oder auch Paratexte vermeiden, die die Frage erweitern.

- **Aufwändige Erstellung:** Die Formulierung zielführender formativer und auf Bildungseffekte angelegter Fragen ist zeitlich aufwändig. Sie können jedes Semester Ihre Fragendatenbank ausbauen. Auch wenn Sie im ersten Jahr z. B. nur rund zwanzig Fragen anbieten, erleichtern Sie dennoch den Wissenserwerb der Studierenden.

6. Vorzüge und Grenzen

Die tabellarische Übersicht über die Vorzüge und Grenzen von MC-Tests sollte vor dem Hintergrund der oben erörterten guten und schlechten Praxisbeispiele gelesen werden. Gewisse Vorzüge werden durch schlecht konzipierte MC-Tests zunichte gemacht, während zielführend konzipierte MC-Tests bestimmte Grenzen aufheben. Es gibt keine alleinseligmachende pädagogische Methode, kein derartiges Instrument und keinen derartigen Ansatz! MC-Tests sind eines von zahlreichen anderen **Instrumenten zur Wissensvermittlung, zum Wissenserwerb und zur Beurteilung von Wissen und Können**. Wir sollten nicht vergessen, dass Pädagogik eine Kunst ist und dass jeder Lehrperson die Freiheit zusteht, sie nach eigener Empfindung, eigenem Fachgebiet und eigenen Neigungen auszuüben. Ihr Einsatz als Lehrperson ist die beste Voraussetzung für den Erfolg Ihrer Studierenden.

Vorzüge	Grenzen
Wenig Korrekturaufwand Geringer Zeitaufwand bei der Korrektur einer grossen Anzahl Tests	Beträchtlicher Zeitaufwand für den Entwurf von zielführenden MC-Tests
Objektive und wahrheitsgetreue Korrektur	Mögliche Zufallstreffer bei exakten Antworten
Evaluation einer breiten Palette von Kompetenzen: Wissen, Verständnis, Analyse, Synthese, Beurteilung.	Verzicht auf die Prüfung der Fähigkeit, komplexe Gedankengänge schriftlich zu formulieren, zu fassen oder vorzubringen.
Der progressive Aufbau einer Fragendatenbank erleichtert die Erstellung zukünftiger MC-Tests.	Geringe Beliebtheit in der Zeugnisphase.
Erkennen der Erwartungen der Lehrperson	In der Zeugnisphase reichen MC-Tests nicht aus, um die Urteilsfähigkeit oder die Fähigkeit zur kritischen Analyse des Stoffs zu evaluieren. Fragmentierung der Kenntnisse.
Abdeckung eines grösseren Wissensspektrums möglich als durch eine einzige Frage.	Die Kandidierenden müssen die Antwort nicht unbedingt wissen, wenn sie es schaffen, die Distraktoren zu eliminieren.
Feedbacks zu Fehlern tragen wirksam zu ihrer Behebung bei und fördern das Erlernen der korrekten Lösungen.	Risiko, dass die Studierenden falsche Lösungen auswendig lernen. Durch den Umgang mit Fehlern steigt das Risiko, dass Fehler erlernt werden.

Wir schliessen unsere Ausführungen mit einem Vergleich zwischen MC-Tests und dem Verfassen schriftlicher Arbeiten (Essay, Fallstudie, Antworten zu einem Beispielfall, offene Fragen, Erörterungen, ...). Studierende, die ihre Antworten entwickeln müssen (AE) nehmen kognitive Kompetenzen in Anspruch, die bei einem MC-Test weniger gefordert sind (Strukturierung der Gedankengänge, Originalität, Induktion, Deduktion, Abduktion, ...). Mit MC-Tests lässt sich dagegen ein weiteres Feld durchkämmen, wenn auch auf Kosten des Tiefgangs. MC-Tests können in kurzer Zeit Dutzende von Themen und ganz unterschiedliche Sichtweisen abdecken, während AE-Tests die Studierenden zwingen, ihre Gedankengänge zu präzisieren, und ihnen Gelegenheit geben, sich auszudrücken.

Dieser Leitfaden soll dazu beitragen, dass Sie (Übungs-)Tests, welche die Entwicklung von Antworten erfordern, nach wie vor ins Zentrum Ihrer Überlegungen stellen. Gleichzeitig sollten Sie sich aber nicht scheuen, im Lauf des Semesters formative, tutoriell begleitete oder diagnostische MC-Tests einzusetzen. Sie wirken sich äusserst positiv auf den Wissenserwerb in einem hybriden universitären Umfeld aus (Spanjers et al., 2015).

7. Literaturverzeichnis

Alle Zitate wurden aus der französischen Version ins Deutsche übersetzt durch Anne Rüsing und Irène Thellung, Synopsis Wirtschaftskommunikation (April 2020).

7.1 Wissenschaftliche oder pädagogische Werke

- Bertrand, M. (2014). *L'utilisation du QCM est-elle un problème ?* PEPS : pour une pédagogie participative et sociale.
<https://pepsagogie.wordpress.com/2014/07/24/lutilisation-du-qcm-est-elle-un-probleme/>
- Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. Washington: NAP.
- Bravard, S. (2004). *Usages pédagogiques des QCM. Un guide pour la mise en place d'un questionnaire à choix multiple*. Diplomarbeit. Universität Poitiers.
<https://docplayer.fr/5833326-Usages-pedagogiques-des-qcm.html>
- Brown, P. C., Roediger, H.L., & McDaniel, M.A. (2014). *Mets-toi ça dans la tête ! Les stratégies d'apprentissage à la lumière des sciences cognitives*. Genf: Markus Haller.
- Dessus, P., Tarbouriech, N. (2019). *Evaluer les apprentissages*. MEEF-PIF : Univ. Grenoble Alpes.
- Johnstone, A. (2003). *Effective practice in objective assessment. The skills of fixed response testing*. LTSN Physical Sciences.
- Krebs, R. (2004). *Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung*. Institut für Medizinische Lehre.
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKewiP-ZbomMzoAhUL6OAKHeA_BrkQFjAAegQIBhAB&url=https%3A%2F%2Fwww.iml.unibe.ch%2Fattachment%2F7%2Fdownload%2Fmc_anleitung.pdf&usq=AOvVaw1FtILT_b85SI2NGrZKFluG
- Prosperi, O. (2015). Le réalisme avec degrés de certitude. *Mesure et évaluation en éducation*, 38 (1), 121–140. <https://doi.org/10.7202/1036553ar>
- Sabatier, P. (2004). Service d'Appui à la Pédagogie, *Guide d'élaboration de questionnaires d'évaluation des connaissances sous forme de QCM*. Université Toulouse III.
http://sup.ups-tlse.fr/documentation/docs/fich_272.pdf und http://sup.ups-tlse.fr/documentation/docs/fich_271.pdf
- Shute, V. J. (2008). *Focus formative feedback*. *Review of Educational Research*, 78(1), SS. 153–189.
- Spanjers, I. A.E., Könings, K. D., Leppink, J., Verstegen, D. M.L., Jong, N. d., Czabanowska, K., & Merriënboer, J. J.G. v. (2015). *The promised land of blended learning: quizzes as a moderator*. *Educational Research Review*, 15, SS. 59–74.
- Sternberg, R.J., Grigorenko, E.L., & Zhang, L. (2008). *Styles of learning and thinking matter in instruction and assesment*, *perspectives on psychological Science* 3, SS. 486-506.
- Unisciel (L'Université des sciences en ligne). *Banque de tests de positionnement. Guide de rédaction des questions d'évaluation*.
<http://unisciel.fr/wp-content/uploads/2017/11/GuideRedactionQCM2018.pdf>

7.2 Ergänzende Bibliographie

- Appleby, D. C. (2008). *A cognitive taxonomy of multiple-choice questions*. In L. T. Benjamin (Ed.), *Favorite activities for the teaching of psychology* (SS. 119-123). Washington, DC: American Psychological Association. (Übersetzung durch Dessus P., 2019).
- Bouillet, D., & Monget, M.-C. (2016). *De l'usage des quiz dans Moodle : retour d'expériences et questions posées*. HAL, archives-ouvertes.
<https://hal.archives-ouvertes.fr/hal-01314850/document>

7.3 Weitere Werke

- Borst, G. zitiert durch Sandrine, C. (23.05.2016). *Les neurosciences peuvent-elles sauver l'école*. Le Monde.
- Montaigne, M. (1580/1595), *Essais* I,26, herausgegeben und übersetzt durch Naya, E., Reguig, D., & Tarrête, A. Paris: Gallimard (2009).
- Quéva, R. (2008). *La science est un jeu*. Paris : Libro.
- Quintilien, M.F. *De l'Institution Oratoire*, herausgegeben und übersetzt durch Cousin, J. (1976/1980), Paris : les belles lettres, (livres II et X).
- Reynaud, F. (1976). *Heureux !* Paris: Gallimard.