Informing the uninformed: a multitier approach to uncover students' misconceptions on cardiovascular physiology

Versteeg, M.; Wijnen-Meijer, M.; Steendijk, P.

HOW WE TEACH | *Generalizable Education Research*

Informing the uninformed: a multitier approach to uncover students' misconceptions on cardiovascular physiology

Marjolein Versteeg,[1,2] **Marjo Wijnen-Meijer,**[3] **and Paul Steendijk**[1,2]

[1]*Center for Innovation in Medical Education, Leiden University Medical Center, Leiden, The Netherlands;* [2]*Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; and* [3]*Medical Education Center, Technical University Munich, Munich, Germany*

Versteeg M, Wijnen-Meijer M, Steendijk P. Informing the uninformed: a multitier approach to uncover students' misconceptions on cardiovascular physiology. *Adv Physiol Educ* 43: 7–14, 2019; doi:10.1152/advan.00130.2018.—Misconceptions about physiology are a major threat to accurate scientific and clinical reasoning in medical education. Awareness is often mentioned as a prerequisite to achieve conceptual understanding; however, students are frequently unaware of their incorrect understanding. We explored the multitier approach as a tool to obtain insight into students' awareness and misconceptions regarding cardiovascular physiology. Biomedical sciences students ($N = 81$) participated in a diagnostic multitier assessment on cardiovascular physiology. Each question consisted of an answer tier and an explanation tier. Both tiers were paired with confidence tiers, i.e., 5-point Likert scales, which were used as an indicator for metacognitive evaluation, i.e., awareness. The average test score was 3.07 (maximum 4.0) for the answer tier only, and reduced to 1.57 when answer and explanation tiers were combined. A weak correlation ($R^2 = 0.13$, $P = 0.001$) between students' confidence and their test scores was found for the combined responses. Correct combined answers were associated with an increase in confidence score of 0.27 vs. incorrect answers. Using a Bland-Altman analysis, we showed that students generally overestimated their knowledge. In total, 28.7% of all responses were classified as misconceptions, defined as incorrect answers paired with high confidence. In all, findings indicate that the multitier approach is useful to study students' conceptual understanding and uncover misconceptions on cardiovascular physiology. Furthermore, this study supports the need for metacognitive measures in order to improve teaching and learning in medical education.

confidence; instructional design; medical education; metacognition; quantitative research methods

INTRODUCTION

The awareness of being correct forms an important component of one's knowledge and ability to learn. This idea was already discussed by early and highly influential philosophers, including Aristotle (c. 300 BC) and Confucius (c. 500 BC), and is still acknowledged today (8, 16, 21, 27). Awareness about one's own thinking and correctness of knowledge after performing a task is here referred to as metacognitive evaluation (16, 35). In medical education, students show a lack of meta-

cognitive evaluation skills consistently across medical training: reduced learning outcomes have been demonstrated for prescribing drugs (4), clinical procedures (28, 31, 32), and evidence-based medicine (26). Remarkably, little emphasis is put on metacognitive evaluation with regard to conceptual understanding of basic sciences, including medical physiology. Nonetheless, accurate recognition of knowing or not knowing something impacts on students' knowledge acquisition (17, 37, 45, 46) and is important for alleviating potential misconceptions in concept learning (7, 36). Any misconceptions on basic science concepts may impair scientific and clinical reasoning, potentially leading to diagnostic errors in medicine (3, 9, 18, 48).

Misconceptions are defined as strongly held ideas that are not in line with current scientific views (7, 41, 47). Since students have limited metacognitive evaluation skills, we need to develop reliable instruments that can amend this issue (24, 38). Making students aware of their level of conceptual understanding may help to alleviate any misconceptions. For instance, research on metacognitive evaluation has shown that students' rating their confidence in their answer could gain deeper insight in their thoughts and potential misbeliefs. Additionally, such confidence ratings may help educators to determine students' actual knowledge (23, 42).

When students remain unaware of their lack of knowledge or misconceptions and subsequently add new information to their current mental structures, this may result in inconsistent thinking (7, 36). Piaget's assimilation theory (34) states that, if there is no fit between the new and the existing information, new knowledge becomes compartmentalized and further strengthens the misconceptions. Educational theorists state that some scientific concepts are difficult to learn because students already hold knowledge that is embedded in naive frameworks, e.g., personal everyday life experiences, and this knowledge is inconsistent with the scientific view (12, 13, 36). Misconceptions are acknowledged to be highly resistant to change (7, 41) and may require educational interventions that differ from the current teaching practice, which only focuses on resolving students' lack of knowledge by providing new information without explicitly addressing misconceptions. Students holding misconceptions may, for instance, benefit from "reshaping" their existing knowledge, also referred to as conceptual change instruction (21), rather than receiving additional factual information or feedback (2, 14, 15). Thus it is of critical importance to design instruments that allow for measuring students' con-

Address for reprint requests and other correspondence: M. Versteeg, Dept. of Cardiology, LUMC, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands (e-mail: m.versteeg@lumc.nl).

ceptual understanding, including assessment of their level of awareness via metacognitive evaluation. By implementing such diagnostic tests in the basic sciences curriculum, potential misconceptions can be detected and alleviated accordingly.

One of the most frequently used forms of assessment in medical education are multiple-choice (MC) questions (1, 29). MC questions allow reliable testing of large cohorts and permit evaluation of higher-order problem-solving (39), but they are not yet widely applied as instruments for detection of misconceptions. Some studies in medical education research have equipped traditional MC questions with a confidence rating scale (10, 11, 19, 23, 37, 46). This allows educators to differentiate between students who are competent (i.e., high confidence-correct), who are guessing (i.e., low confidence-correct), who have a lack of knowledge (i.e., low confidence-incorrect), and who have a potential misconception (i.e., high confidence-incorrect). These MC questions paired with confidence scales are a first attempt toward creating awareness and subsequently uncovering misconceptions by using metacognitive measures in assessments. In the domain of science education, Treagust and colleagues (43, 44) took the MC assessment a step further. They have developed a two-tier diagnostic test to pinpoint students' potential misunderstanding of the subject matter. In the answer tier, students have to make a binary choice (yes/no, higher/lower) about some specific content knowledge. In the explanation tier, students are asked to mark a reason or explanation that supports their choice in the first tier. In the two-tier format, however, it is difficult to distinguish between an accurate understanding or guesswork. Moreover, for incorrect answers, one cannot determine whether the response is a consequence of a lack of knowledge or due to a misconception. As outlined above, the confidence ratings may allow for these classifications. Combining such a two-tier test with additional confidence ratings is also referred to as a multitier approach, which has shown its potential in science education. The multitier approach has been applied in various formats, including the four-tier format, which includes separate confidence scales for each of the two tiers (5, 40, 49).

The multitier approach is a promising tool to measure conceptual understanding; however, its effectiveness in medical education remains to be explored. The use of a two-tier diagnostic test has been reported in the medical field already (33), but has not yet been paired with confidence ratings. We aim to investigate if a multitier approach provides information about students' conceptual understanding and potential misconceptions regarding cardiovascular physiology.

The implementation of the multitier approach in medical education may provide insight in students' conceptual understanding and distinguish cases with a potential lack of knowledge from those who hold strong misconceptions. This is useful feedback that can be used by both students and their educators to improve learning and teaching, respectively. Since basic science knowledge forms an important foundation for scientific and clinical reasoning (3, 9, 18, 48), we investigate the use of multitier assessments in the context of basic cardiovascular physiology concepts. This research is set out to address the following questions. 1) Can a multitier approach provide information on students' conceptual understanding by assessing their metacognitive evaluation skills? 2) What are the prominent misconceptions regarding cardiovascular physiology and their prevalence among biomedical students?

MATERIALS AND METHODS

Participants. Eighty-one biomedical sciences students voluntarily participated in this experimental study. These were second-year bachelor students enrolled in a "Physiology Basic Concepts" course. The male-to-female ratio in this cohort was 30:70.

Ethical considerations. The educational research study was announced at the beginning of the course, and before the test students were asked to provide informed consent to use their anonymous answers for educational research. They could withdraw their permission at any time. Students received no additional credit, and they were informed that test performance had no effect on their course grade. The study protocol was reviewed and approved by the Leiden University Medical Center Institutional Scientific Committee on Educational Research.

Study design. As part of the course setup, students were enrolled in seminar groups (~20 students/group). During the seminar, students performed a test consisting of four questions concerning basic cardiovascular physiology concepts: 1) systolic heart failure and ejection fraction; 2) cardiac output and mean arterial blood pressure; 3) transit time in pulmonary and systemic circulations; and 4) afterload and stroke volume. Each question consisted of four tiers, i.e., an answer tier, an explanation tier, and two confidence tiers (Fig 1). In the answer tier, students were asked to provide a binary yes/no (Y/N) answer. In the explanation tier, students could choose one of the suggested explanations (4-6 options) that best supported the reasoning underlying their Y/N answer. Note that each option could be a correct statement in and of itself, but that the students should choose the option that best explains their given Y/N answer. The questions and explanations were designed by a physiology teacher (P.S.) with longstanding experience in cardiovascular research and teaching and designing and reviewing exam questions. We selected four topics that were handled in the course and on which, based on our experience, misconceptions are relatively common. We aimed for concepts that could be tested by compact statements for which relatively short correct explanations and multiple "plausible" alternative incorrect explanations could be formulated.

Confidence was assessed on both the answer and the explanation: "How sure are you that your answer to the previous question was correct?" Confidence was self-graded using a 5-point Likert scale: 1 = very unsure (complete guess), 2 = fairly unsure, 3 = in doubt, 4 = fairly sure, 5 = very sure (almost 100%).

Outcome measures. To determine whether a multitier approach can make students aware of their conceptual understanding by assessing metacognitive evaluation, we used various outcome measures. First, we reported the performance scores and corresponding confidence scores for each question. Performance scores are also given corrected for guessing (e.g. 25% guess chance for a four-option MCQ) using the following formula: (score – guess score)/(maximum score – guess score). Second, we computed the correlation between overall test performance (i.e., actual knowledge) and average confidence level (i.e., self-perceived knowledge) for each student, using Spearman's rank correlation coefficient.

Third, to determine the within-student difference in confidence between incorrect and correct answers, confidence scores were corrected for each student's average confidence score, and subsequently the difference in confidence per question was determined by linear regression.

Following the idea that knowledge is not solely a matter of being able to provide correct answers (performance) but also incorporates students' confidence in it, Kampmeyer and colleagues (23) introduced the knowledge value. The knowledge value combines correctness and confidence in a single value and allows one to distinguish between learners who have similar test scores, but who differ in their metacognitive evaluation. For example, a student who has the maximum test score but who underestimated her/himself (i.e., a low metacognitive evaluation score) has a lower knowledge value than a student

Question

The transit time of a red blood cell through the pulmonary circulation is less than its transit time through the systemic circulation.

(1) Answer tier

Is this statement correct?

> A. Yes*
> B. No

(2) Confidence tier

How sure (confident) are you that your answer to the previous question was correct?

> A. Very unsure (complete guess)
> B. Fairly unsure
> C. In doubt (50/50)
> D. Fairly sure
> E. Very sure (almost 100%)

(3) Explanation tier

Select an explanation for your answer: Because the pulmonary circulation and the systemic circulation are connected in series and ...

> A. Pulmonary blood volume is less than systemic blood volume*
> B. Pulmonary vascular resistance is less than systemic vascular resistance
> C. Right ventricular pressure is lower than left ventricular pressure
> D. Blood viscosity is the same in both systems
> E. Mean blood velocity is the same in both systems
> F. The flow is the same in both systems

(4) Confidence tier

How sure (confident) are you that your answer to the previous question was correct?

> A. Very unsure (complete guess)
> B. Fairly unsure
> C. In doubt (50/50)
> D. Fairly sure
> E. Very sure (almost 100%)

Fig. 1. Example of a multitier question derived from our concept test. *Correct answers.

who has both high test and metacognitive evaluation performance. For this analysis, correctness and confidence were both normalized on a 0–1 scale, and we defined knowledge value = $0.5 \times$ (confidence + correctness) and centration = (confidence – correctness) using the Bland-Altman analysis.

We used Hasan's decision matrix to examine the prevalence of misconceptions among students (20). To this end, confidence levels were reduced to dichotomous outcomes: *levels 1*, *2*, and *3* were scored as low confidence; *levels 4* and *5* as high confidence. This cut-off was chosen because students selecting "3" were still essentially unsure ('in doubt') about the correct answer. If the student provided the wrong response to the answer tier and indicated that he/she was fairly sure (*level 4*) or very sure (*level 5*), a misconception was assumed to be present.

Via the explanation tier, we further tested the students by asking for the underlying arguments. If the student failed to provide the correct explanation, this was also taken to indicate a misconception, even if the Y/N answer was correct.

Table 1. *Students' correctness and associated confidence on the multitier instrument*

Question	Answer Tier Y/N			Explanation Tier			Combined Tiers		
	Correct, %	Confidence	SD	Correct	Confidence	SD	Correct	Confidence	SD
Q1	91.5 (83.0)	3.77	0.74	54.9 (39.9)	3.16	0.84	54.9 (48.5)	3.49	0.59
Q2	68.3 (36.6)	3.38	0.77	26.8 (12.1)	2.96	0.94	23.2 (16.2)	3.17	0.67
Q3	75.6 (51.2)	3.67	0.85	30.5 (16.6)	3.06	0.86	29.3 (22.9)	3.36	0.73
Q4	70.7 (41.4)	3.81	0.84	46.3 (35.6)	3.48	1.01	45.1 (40.1)	3.65	0.84

$N = 81$ students. Scores in parentheses are adjusted for guessing. Y/N, yes/no.
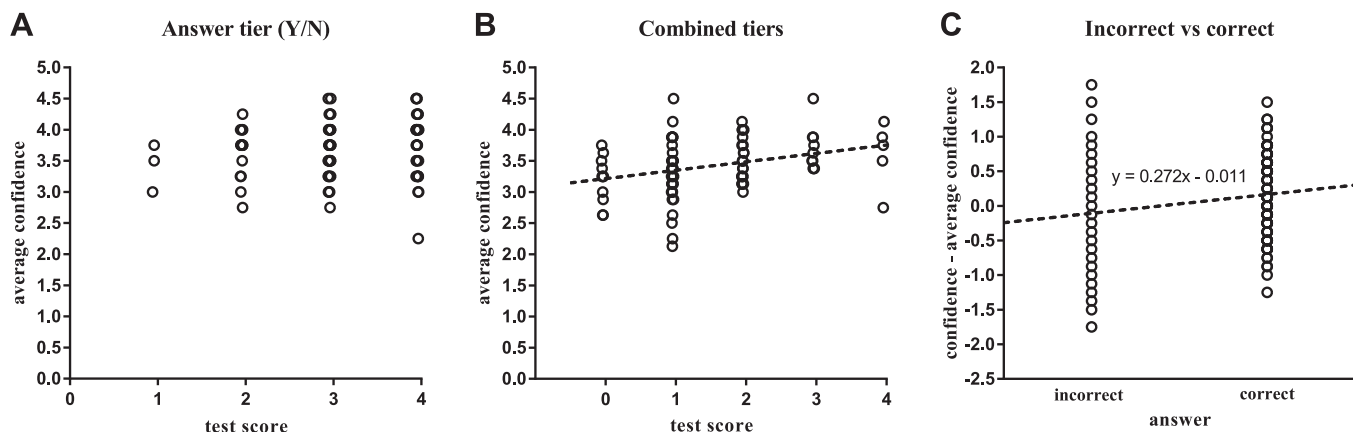
Fig. 2. Relations between confidence and correctness. Average confidence vs. test score for the answer tier (*A*) and for the combined tiers (*B*) is shown. *C*: regression analysis (within students) shows an increase in confidence related to the correctness of answers. Y/N, yes/no.

To identify the misconceptions more specifically, we adopted a quantitative analysis proposed by Caleon and Subramaniam (6). To obtain robust results, we classified only incorrect answers and explanations that were chosen by at least 10% of the participants as real alternative conceptions. Accordingly, we calculated the associated confidence in these alternative conceptions (CAC), the average confidence rating of students with this conception. Since we used a cut-off of 3, spurious alternative conceptions in this study have a CAC value between 3 and 3.5 and strong alternative conceptions yield a CAC >3.5.

RESULTS

Metacognitive evaluation. A total of 91.5% students provided a correct response on the answer tier of *question 1* (Table 1). Correcting for the 50% guess score, this results in a score of 83.0%. The associated average confidence score was 3.77 out of 5. The explanation tier of *question 1* was answered correctly by 54.9% of all students, paired with an average confidence of 3.12. When combining the answers, 54.9% provided a correct answer on both the answer and explanation tier, meaning that almost 40.0% of initially correct students failed to mark the correct explanation. The average confidence

for the combined tiers was 3.45. Similar results were obtained for the other questions (Table 1).

The average total test scores (maximum 4 points) among students when combining the answer and explanation tier was 1.57, compared with 3.07 for the answer tier only, yielding a significant reduction in performance [$t(80) = 13.209$, $P < 0.0001$, $d = 1.56$]. The self-assessed confidence levels were also significantly lower for the combined tiers (3.42) vs. the answer tier responses (3.67) [$t(80) = 9.337$, $P < 0.0001$, $d = 0.55$]. The average confidence levels for the answer tier responses and for the combined responses were both above 3 for all questions.

For the answer tier, there was no significant correlation between students' test scores and their average level of confidence (Fig 2A). For the combined tiers, a weak positive correlation [$R^2 = 0.13$, $P = 0.001$] was found (Fig 2B). The average confidence for a correct response on a Y/N question was 3.72 vs. 3.34 for an incorrect response [$t(322) = 1.940$, $P = 0.053$]. For the combined tiers, the average confidence was 3.67 for correct answers and 3.21 for incorrect answers [$t(322) = 2.711$, $P = 0.007$]. After removing the between-
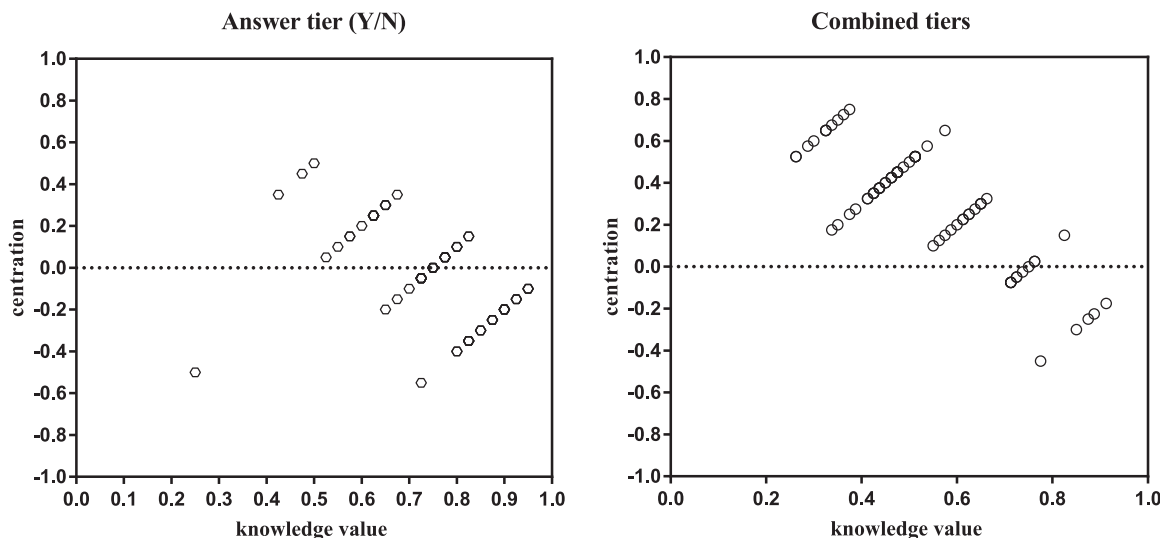

Fig 3. Bland-Altman analysis. The knowledge values [$0.5 \times$ (confidence + correctness)] are plotted against the mean centration values (confidence − correctness) for each individual student. Y/N, yes/no.

student variability in average confidence, the specific effect of an incorrect vs. a correct combined answer was determined as 0.27 on the 5-point confidence scale (Fig 2C).

In addition to the correlation analysis, we performed a Bland-Altman analysis to relate students' knowledge values and centration values (compare MATERIALS AND METHODS, Fig 3). The average knowledge value was higher for the answer tier (0.74) compared with the combined tiers (0.54). The mean centration is negative for Y/N answers (−0.03), which indicates a bias between actual knowledge and confidence with a tendency toward underestimation for Y/N responses. For the combined tiers, the mean centration is positive (0.29), meaning that, based on their multitier assessment, students, on average, overestimate their actual knowledge.

Misconceptions. The prevalence of misconceptions was computed using Hasan's decision matrix (Fig 4). Looking only at the answer tiers of each question, 10.4% of all answers were categorized as misconceptions. Almost one-half of the Y/N answers were answered correctly with high confidence (48.8%). Moreover, 27.7% of the answers were correct but paired with low confidence (i.e., lucky guesses), and 13.1% were recognized as a lack of knowledge.

If also the responses on the explanation tier are taken into account, the distribution of outcomes clearly changes. The percentage of misconceptions nearly triples toward 28.7%, and

Answer tier (Y/N)

	Incorrect	Correct
High confidence	10.4%	48.8%
Low confidence	13.1%	27.7%

Combined tiers

	Incorrect	Correct
High confidence	28.7%	29.9%
Low confidence	32.6%	8.8%

Fig. 4. Hasan's decision matrix. Y/N, yes/no.

only 29.9% of all answers are categorized as right conceptions. The percentage of lucky guesses reduces to 8.8%, and lack of knowledge increases to 32.6%.

Table 2 shows a list of seven misconceptions on cardiovascular physiology that were identified in our study. The mean confidence for these misconceptions ranges between 2.97 and 4.00. Using the classification scheme of Caleon and Subramanian (6), we classified two of these responses (Q3$_{M2}$ and Q4$_{M1}$) as strong alternative conceptions.

DISCUSSION

Our study shows that the multitier approach can provide information about students' level of conceptual understanding and their associated metacognitive evaluation skills. Interestingly, students frequently chose an incorrect explanation they believed was associated with their initial correct Y/N response. These incorrect explanations were often paired with high confidence; therefore, we conclude that misconceptions are clearly present among students enrolled in this cardiovascular physiology course.

Metacognitive evaluation. Based on the Y/N responses, students yielded relatively high performance scores on the test. Students performed significantly worse when they had to mark the explanation they believed was associated with their Y/N response (Table 1). Remarkably, although significantly lower, the self-reported confidence levels were not altered substantially on the multitier assessment compared with the confidence in students' Y/N responses. Average confidence responses were still between "in-doubt" and "fairly sure" and indicated an overestimation of students' knowledge when considering the combined tiers. Furthermore, the relation between students' actual knowledge and confidence showed only a weak correlation (Fig 2). These results are in line with previous literature demonstrating a tendency of students to overestimate themselves (4, 25, 26). Using the Bland-Altman approach, we confirmed that students indeed overestimate their actual knowledge, primarily when they have to choose the right explanation for a question (Fig 3). The negative correlation between the knowledge values and the centration indicated that the overestimation is less for students with a higher knowledge value. These results are in contrast with previous findings from Kampmeyer and colleagues (23), who found a relatively low percentage of incorrect high-confident answers. However, Kampmeyer et al. used traditional MC questions instead of a multitier approach, which complicates the interpretation of differences in study outcomes.

The difference in confidence responses between correct and incorrect answers is significant in our study, but only for the combined tiers. Our findings are supported by previous studies, which have shown that students' confidence in correct responses is higher (23, 37).

Misconceptions. The prevalence of misconceptions was 10.4%, considering Y/N responses only, and increased to 28.7% when including the explanation tier (Fig 4). The number of incorrect answers paired with high confidence (i.e., misconceptions) was almost the same as those paired with low confidence (i.e., lack of knowledge), indicating that educators should equally focus on both categories. Notably, the percentage of misconceptions in our study was lower than reported previously (30, 33), although these studies did not include

Table 2. *Misconceptions on cardiovascular physiology*

Question	Misconception	%Students with Misconception	CAC Mean	SD
Q1	A patient with systolic heart failure will have a low ejection fraction, because with systolic heart failure *end-diastolic volume and stroke volume are both decreased*.	34.6	3.46	0.70
Q2$_{M1}$	During exercise the %increase in cardiac output is *not* approximately the same as the %increase in mean arterial blood pressure. Because during exercise *the increase in cardiac output is mainly due to an increase in heart rate*.	30.9	3.12	0.60
Q2$_{M2}$	During exercise, the %increase in cardiac output is approximately the same as the %increase in mean arterial blood pressure. Because during exercise *the increase in cardiac output is mainly due to an increase in heart rate*.	21.0	2.97	0.65
Q3$_{M1}$	The transit time of a red blood cell through the pulmonary circulation is *not* less than its transit time through the systemic circulation. Because the pulmonary circulation and the systemic circulation are connected in series and *pulmonary vascular resistance is less than systemic vascular resistance*.	28.4	3.22	1.03
Q3$_{M2}$	The transit time of a red blood cell through the pulmonary circulation is less than its transit time through the systemic circulation. Because the pulmonary circulation and the systemic circulation are connected in series and *the flow is the same in both systems*.	11.1	3.61	0.97
Q4$_{M1}$	An increase in afterload will generally cause a decrease in stroke volume. Because with an increase in afterload *end-systolic volume will decrease*.	16.0	4.00	0.94
Q4$_{M2}$	An increase in afterload will generally *not* cause a decrease in stroke volume. Because with an increase in afterload *end-systolic volume will decrease*.	11.1	3.11	1.38

CAC, confidence for alternative conceptions.

confidence measures. For cardiovascular physiology, the most prevalent misconceptions of our test were outlined (Table 2). For some questions, the average CAC was higher than for the correct answer. In all, findings indicate that students' basic science knowledge on cardiovascular physiology is insufficient. Therefore, educators should not only design conceptual change interventions to alleviate the prevalence of misconceptions, but also focus on the apparent lack of knowledge among students by examining their prior knowledge, for example.

Probability of guessing. The difference between high scores on the answer tier vs. relatively low scores on the combined tiers may be partly explained by the probability of guessing. Since the answer tier is associated with a high chance of guessing (50%), students' scores will be an overestimation of their actual knowledge. The difference between Y/N and combined tiers might thus be more subtle. Therefore, we applied a correction for the probability of guessing to the absolute values (Table 1). Although the difference is smaller after correction for guessing, it remains significant.

Strengths and limitations. To our knowledge we are the first to evaluate the use of the multitier approach in medical education. We used the multitier approach to uncover students' conceptual understanding. Additionally, we showed that there is a clear lack of metacognitive evaluation skills among students regarding basic science knowledge. Still, this study contains several limitations that should be addressed. First, our findings are based on a relatively small set of questions testing conceptual understanding related to cardiovascular physiology. It would be interesting to investigate the multitier approach in other contexts, such as other basic science knowledge or clinical skills. Second, we limited ourselves to exploring the single relationship between

students' confidence and performance. Other factors, such as motivation and question type, may have a substantial influence on both confidence ratings and performance and should be analyzed in future studies. Third, we used a 5-point confidence scale for both tiers, although the answer tier only had two answer options. It may not seem logical for students to give a rating <3, as this rating was defined as doubting between two answer options. However, no students commented on this issue, and remarkably 5.9% of the confidence ratings associated with the Y/N response were <3. This finding illustrates that more methodological research might be needed on students' interpretation of confidence rating scales. Lastly, this experiment did not provide a representative overview of existing misconceptions in cardiovascular physiology, since the assessment only comprised four questions. Instead, this study is considered a proof-of-concept study demonstrating that the multitier approach is useful for detecting students misconceptions in medical education.

Conclusion. We showed that the multitier approach allows students and their educators to gain insight in students' level of conceptual understanding and to reveal their potential misconceptions. Broad implementation of the multitier diagnostic test can help educators to more precisely pin-point knowledge deficiencies, which may result in more effective teaching approaches and learning across the medical curriculum.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

M.V. and P.S. performed experiments; M.V. and P.S. analyzed data; M.V., M.W.-M., and P.S. interpreted results of experiments; M.V. and P.S. prepared figures; M.V., M.W.-M., and P.S. drafted manuscript; M.V., M.W.-M., and P.S. edited and revised manuscript; M.V., M.W.-M., and P.S. approved final version of manuscript.

REFERENCES

1. **Anderson J.** For multiple choice questions. *Med Teach* 1: 37–42, 1979. doi:10.3109/01421597909010580.
2. **Archer JC.** State of the science in health professional education: effective feedback. *Med Educ* 44: 101–108, 2010. doi:10.1111/j.1365-2923.2009.03546.x.
3. **Berner ES, Graber ML.** Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 121, Suppl: S2–S23, 2008. doi:10.1016/j.amjmed.2008.01.001.
4. **Brinkman DJ, Tichelaar J, van Agtmael MA, de Vries TP, Richir MC.** Self-reported confidence in prescribing skills correlates poorly with assessed competence in fourth-year medical students. *J Clin Pharmacol* 55: 825–830, 2015. doi:10.1002/jcph.474.
5. **Caleon IS, Subramaniam R.** Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Res Sci Educ* 40: 313–337, 2010. doi:10.1007/s11165-009-9122-4.
6. **Caleon I, Subramaniam R.** Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *Int J Sci Educ* 32: 939–961, 2010. doi:10.1080/09500690902890130.
7. **Chi MT, Roscoe RD, Slotta JD, Roy M, Chase CC.** Misconceived causal explanations for emergent processes. *Cogn Sci* 36: 1–61, 2012. doi:10.1111/j.1551-6709.2011.01207.x.
8. **Colman AM.** *The Dictionary of Psychology.* Oxford, UK: Oxford University Press, 2001.
9. **Croskerry P, Norman G.** Overconfidence in clinical decision making. *Am J Med* 121, Suppl: S24–S29, 2008. doi:10.1016/j.amjmed.2008.02.001.
10. **Curtis DA, Lind SL, Boscardin CK, Dellinges M.** Does student confidence on multiple-choice question assessments provide useful information? *Med Educ* 47: 578–584, 2013. doi:10.1111/medu.12147.
11. **Curtis DA, Lind SL, Dellinges M, Schroeder K.** Identifying student misconceptions in biomedical course assessments in dental education. *J Dent Educ* 76: 1183–1194, 2012.
12. **Duit R, Treagust DF.** How can conceptual change contribute to theory and practice in science education? In: *Second International Handbook of Science Education,* edited by Fraser B, Tobin K, McRobbie C. Dordrecht, Germany: Springer, 2012, p. 107–118. doi:10.1007/978-1-4020-9041-7_9.
13. **Duit R, Treagust DF.** Students' conceptions and constructivist teaching approaches. In: *Improving Science Education,* edited by Fraser BJ, Walberg HJ. Chicago, IL: University of Chicago Press, 1995, p. 46–69.
14. **Ecker UK, Lewandowsky S, Swire B, Chang D.** Correcting false information in memory: manipulating the strength of misinformation encoding and its retraction. *Psychon Bull Rev* 18: 570–578, 2011. doi:10.3758/s13423-011-0065-1.
15. **Ecker UK, Lewandowsky S, Tang DT.** Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Mem Cognit* 38: 1087–1100, 2010. doi:10.3758/MC.38.8.1087.
16. **Flavell JH.** Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am Psychol* 34: 906–911, 1979. doi:10.1037/0003-066X.34.10.906.
17. **Friedman CP, Gatti GG, Franz TM, Murphy GC, Wolf FM, Heckerling PS, Fine PL, Miller TM, Elstein AS.** Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *J Gen Intern Med* 20: 334–339, 2005. doi:10.1111/j.1525-1497.2005.30145.x.
18. **Grande JP.** Training of physicians for the twenty-first century: role of the basic sciences. *Med Teach* 31: 802–806, 2009. doi:10.1080/01421590903137049.
19. **Grazziotin-Soares R, Lind SL, Ardenghi DM, Curtis DA.** Misconceptions amongst dental students: How can they be identified? *Eur J Dent Educ* 22: e101–e106, 2018. doi:10.1111/eje.12264.
20. **Hasan S, Bagayoko D, Kelley EL.** Misconceptions and the certainty of response index (CRI). *Phys Educ* 34: 294–299, 1999. doi:10.1088/0031-9120/34/5/304.
21. **Hewson MG, Hewson PW.** Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *J Res Sci Teach* 20: 731–743, 1983. doi:10.1002/tea.3660200804.
22. **Hunt DP.** The concept of knowledge and how to measure it. *J Intellect Cap* 4: 100–113, 2003. doi:10.1108/14691930310455414.
23. **Kampmeyer D, Matthes J, Herzig S.** Lucky guess or knowledge: a cross-sectional study using the Bland and Altman analysis to compare confidence-based testing of pharmacological knowledge in 3rd and 5th year medical students. *Adv Health Sci Educ Theory Pract* 20: 431–440, 2015. doi:10.1007/s10459-014-9537-1.
24. **Klymkowsky MW, Taylor LB, Spindler SR, Garvin-Doxas RK.** Two-dimensional, implicit confidence tests as a tool for recognizing student misconceptions. *J Coll Sci Teach* 36: 44, 2006.
25. **Kruger J, Dunning D.** Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 77: 1121–1134, 1999. doi:10.1037/0022-3514.77.6.1121.
26. **Lai NM, Teng CL.** Self-perceived competence correlates poorly with objectively measured competence in evidence based medicine among medical students. *BMC Med Educ* 11: 25, 2011. doi:10.1186/1472-6920-11-25.
27. **Legge J.** *Confucian Analects, The Great Learning and The Doctrine of the Mean.* New York: Dover, 1971.
28. **Mavis B.** Self-efficacy and OSCE performance among second year medical students. *Adv Health Sci Educ Theory Pract* 6: 93–102, 2001. doi:10.1023/A:1011404132508.
29. **McCoubrie P.** Improving the fairness of multiple-choice questions: a literature review. *Med Teach* 26: 709–712, 2004. doi:10.1080/01421590400013495.
30. **Michael JA, Wenderoth MP, Modell HI, Cliff W, Horwitz B, McHale P, Richardson D, Silverthorn D, Williams S, Whitescarver S.** Undergraduates' understanding of cardiovascular phenomena. *Adv Physiol Educ* 26: 72–84, 2002. doi:10.1152/advan.00002.2002.
31. **Morgan PJ, Cleave-Hogg D.** Comparison between medical students' experience, confidence and competence. *Med Educ* 36: 534–539, 2002. doi:10.1046/j.1365-2923.2002.01228.x.
32. **Morton J, Anderson L, Frame F, Moyes J, Cameron H.** Back to the future: teaching medical students clinical procedures. *Med Teach* 28: 723–728, 2006. doi:10.1080/01421590601110025.
33. **Palizvan MR, Nejad MR, Jand A, Rafeie M.** Cardiovascular physiology misconceptions and the potential of cardiovascular physiology teaching to alleviate these. *Med Teach* 35: 454–458, 2013. doi:10.3109/0142159X.2013.774331.
34. **Piaget J.** Piaget's theory. In: *Piaget and His School,* edited by Inhelder B, Chipman HH, Zwingmann C. New York: Springer, 1976, p. 11–23.
35. **Pintrich PR.** The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Pract* 41: 219–225, 2002. doi:10.1207/s15430421tip4104_3.
36. **Posner GJ, Strike KA, Hewson PW, Gertzog WA.** Accommodation of a scientific conception: Toward a theory of conceptual change. *Sci Educ* 66: 211–227, 1982. doi:10.1002/sce.3730660207.
37. **Rangel RH, Möller L, Sitter H, Stibane T, Strzelczyk A.** Sure, or unsure? Measuring students' confidence and the potential impact on patient safety in multiple-choice questions. *Med Teach* 39: 1189–1194, 2017. doi:10.1080/0142159X.2017.1362103.
38. **Rovick AA, Michael JA, Modell HI, Bruce DS, Horwitz B, Adamson T, Richardson DR, Silverthorn DU, Whitescarver SA.** How accurate are our assumptions about our students' background knowledge? *Am J Physiol* 276: S93–S101, 1999. doi:10.1152/advances.1999.276.6.S93.
39. **Schuwirth LW, van der Vleuten CP.** Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 38: 974–979, 2004. doi:10.1111/j.1365-2929.2004.01916.x.
40. **Sreenivasulu B, Subramaniam R.** Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures. *Res Sci Educ* 44: 801–828, 2014. doi:10.1007/s11165-014-9400-7.
41. **Taylor AK, Kowalski P.** Naive psychological science: the prevalence, strength, and sources of misconceptions. *Psychol Rec* 54: 15–25, 2004. doi:10.1007/BF03395459.
42. **Treagust D.** Evaluating students' misconceptions by means of diagnostic multiple choice items. *Res Sci Educ* 16: 199–207, 1986. doi:10.1007/BF02356835.

43. **Treagust DF.** Development and use of diagnostic tests to evaluate students' misconceptions in science. *Int J Sci Educ* 10: 159–169, 1988. doi:10.1080/0950069880100204.

44. **Tsui CY, Treagust D.** Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *Int J Sci Educ* 32: 1073–1098, 2010. doi:10.1080/09500690902951429.

45. **Tweed M, Purdie G, Wilkinson T.** Low performing students have insightfulness when they reflect-in-action. *Med Educ* 51: 316–323, 2017. doi:10.1111/medu.13206.

46. **Tweed MJ, Stein S, Wilkinson TJ, Purdie G, Smith J.** Certainty and safe consequence responses provide additional information from multiple choice question assessments. *BMC Med Educ* 17: 106, 2017. doi:10.1186/s12909-017-0942-z.

47. **Wandersee JH, Mintzes JJ, Novak JD.** Research on alternative conceptions in science. In: *Handbook Of Research On Science Teaching And Learning*, edited by Gabel DL. New York: Macmillan, 1994, p. 177–210.

48. **Woods NN.** Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ* 41: 1173–1177, 2007. doi:10.1111/j.1365-2923.2007.02911.x.

49. **Yan YK, Subramaniam R.** Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chem Educ Res Pract* 19: 213–226, 2018. doi:10.1039/C7RP00143F.