"Title Page"

Title: The effect of assessments on student motivation for learning and its outcomes in health professions education: A review

Authors: Rashmi A Kusurkar [1,2,3], Cesar Orsini*[4,5], Sunia Somra*[6], Anthony R Artino Jr.[7], Hester E M Daelmans[8], Linda J Schoonmade[9], Cees van der Vleuten[10]

*Cesar Orsini and *Sunia Somra have contributed equally and are, therefore, joint second authors on this manuscript.

Author Affiliations:

[1]Amsterdam UMC location Vrije Universiteit Amsterdam, Research in Education, De Boelelaan 1118, Amsterdam, The Netherlands

[2]LEARN! Research Institute for Learning and Education, Faculty of Psychology and Education, VU University Amsterdam, The Netherlands

[3]Amsterdam Public Health, Quality of Care, Amsterdam, The Netherlands

[4]Norwich Medical School, University of East Anglia, Norwich, UK

[5]Faculty of Dentistry, Universidad de Los Andes, Chile

[6]GGD, The Hague, The Netherlands

[7]School of Medicine & Health Sciences, The George Washington University, USA

[8]Faculty of Medicine Vrije Universiteit Amsterdam, The Netherlands

[9]Medical Library, Vrije Universiteit Amsterdam, the Netherlands

[10]School of Health Professions Education, University of Maastricht, the Netherlands

Author bylines:

Rashmi A Kusurkar, MD, PhD, is a Professor and Research Programme Leader at Amsterdam UMC location Vrije Universiteit Amsterdam, Research in Education, Amsterdam, the Netherlands. ORCID: http://orcid.org/0000-0002-9382-0379

*Cesar Orsini, MEd, DHPE, is an Associate Professor in Medical Education at Norwich Medical School, University of East Anglia, Norwich, UK, and Researcher in Health Professions Education at the Faculty of Dentistry, Universidad de Los Andes, Chile. ORCID: http://orcid.org/0000000252263625

*Sunia Somra, MSc, was a Research Assistant at Research in Education, Amsterdam UMC Faculty of Medicine, Vrije Universiteit, Amsterdam, the Netherlands, when this study was conducted.

Anthony R Artino Jr., PhD, is a Professor and Associate Dean for Evaluation and Educational Research at School of Medicine & Health Sciences, The George Washington University, USA. ORCID: http://orcid.org/0000-0003-2661-7853

Hester E. M. Daelmans, MD, PhD, is Director of Master of Medicine Programme at the Faculty of Medicine Vrije Universiteit Amsterdam, the Netherlands.

Linda J Schoonmade, is an information specialist at the Medical Library, Vrije Universiteit Amsterdam, the Netherlands. ORCID: https://orcid.org/0000-0002-2407-5977

Cees van der Vleuten, PhD, is Professor at the School of Health Professions Education, University of Maastricht, the Netherlands. ORCID: http://orcid.org/0000-0001-6802-3119

Corresponding author:

Dr. R. A. Kusurkar

Amsterdam UMC location Vrije Universiteit Amsterdam,

PK KTC 7.028,

P.O. Box 7057, 1007 MB Amsterdam,

The Netherlands.

Telephone: +3120 4445790

E-mail: R.Kusurkar@amsterdamumc.nl

Twitter handle: @r_kusurkar

Authors' contributions:

RAK conceptualized and designed the study. All authors, CO, SS, AA, HD, LJS and CvdV provided feedback on the original concept and design of the study. RAK and LS designed the search strategy, LJS implemented the search and contributed to the methods. RAK, CO and SS handled the inclusion-exclusion of the papers in this review and also extracted the data. AA checked the extracted data from the randomly chosen included papers. All authors contributed to the final results, interpretation of the results and the discussion. All authors contributed to and approved the submitted manuscript. All authors agree to be accountable for all aspects of the work.

Acknowledgements:

The authors would like to thank Joyce Kors, M.Sc., Dr. Anouk Wouters, Malou Stoffels, M.Sc., Jettie Vreugdenhil, M.Sc, and Lianne Mulder, M.A. & Master of Philosophy, from Amsterdam UMC, Faculty of Medicine Vrije Universiteit Amsterdam, the Netherlands, and Dr. Andries Koster from University of Utrecht, the Netherlands, for their feedback on an earlier version of this manuscript.

Funding information: None

Ethical approval: Not Applicable

“Revised Manuscript”

Title: The effect of assessments on student motivation for learning and its outcomes in health professions education: A review

Abstract

Purpose

In health professions education (HPE) the effects of assessments on learning have been studied, but their effect on the quality of student motivation for learning and its consequences have been neglected. This is problematic because assessments can hamper motivation and psychological well-being. This review addressed the following research questions: *How do assessments affect student motivation for learning in HPE? What outcomes does this lead to in which contexts?*

Method

The authors searched PubMed, EMBASE, PsycInfo, ERIC, Cinahl and Web of Science for “assessments” AND “motivation” AND “health professions education/students”. Empirical papers or reviews investigating the effect of assessments on student motivation for learning and its outcomes in HPE using quantitative, qualitative, and mixed methods were included. The authors chose the realist synthesis method for data analysis in order to study complex phenomena having context-dependent intended and unintended consequences. Two researchers independently screened the articles, extracted the data, classified assessments as stimulating autonomous/controlled motivation, and determined the context, mechanism and outcomes for each.

Results

The assessments employed in the 24 included articles seemed to stimulate controlled rather than autonomous motivation. An example of assessments stimulating controlled motivation included the following: the assessment had a focus on factual knowledge *(context)*, which through the *mechanism* of encouraging students to study to pass led to strategic learning *(outcome)*. An example of assessments stimulating autonomous motivation included the following: the assessment was fun *(context)*, which through the *mechanism* of active learning led to higher study effort *(outcome)*.

Conclusion

Students strategically learned what was expected in the assessments at the expense of what was needed in professional practice. These findings suggest that health professions educators should rethink their assessment philosophy and practices and introduce assessments that stimulate interest in the content and professional practice.

The effect of assessments on student motivation for learning and its outcomes in health professions education: A review

In higher education, in general, and in health professions education (HPE), more specifically, authors have debated the effects of assessments on learning, with phrases like "Assessment-drives-learning", "Assessment-for/of-learning" and "Assessment-as-learning" pervading the literature.[1-3] However, the effect of assessment on the quality of student motivation and hence on learning outcomes has been largely neglected in this scientific dialogue. This gap is important as high-stakes assessments can not only hamper students' autonomous motivation as a long-term effect,[4] but can also produce psychological distress.[5] Thus, its effect on motivation could be one causal mechanism by which assessment influences learning and psychological well-being.[6,7] Here, psychological well-being includes feeling good and functioning effectively.[7] A compromise on any one or both of these aspects would be characterized as negative psychological well-being.[7] With learning we mean, "an enduring change in behavior or the capacity to behave in a given fashion, which results from practice or other forms of experience".[8] The current review aims to improve our understanding of how assessments influence student motivation which in turn has an effect on learning and psychological well-being.

Along with providing summative evaluation of students' knowledge and skills, educators often intend for assessments to produce learning. Students, on the contrary, often focus on "giving a performance" rather than on learning.[9,10] This represents a major gap in the conceptualization and impact of assessments. The 2018 Consensus framework for good assessment recommends seven criteria including educational and catalytic effects that are concerned with student motivation.[11] Educational effects refer to assessments motivating students to prepare and produce educational benefit, whereas catalytic effects refer to assessments providing results and feedback that motivates stakeholders in creating, improving, and supporting education.[11] In this framework, the concept of motivation is rather limited, as it focuses on the "educational benefit" part whilst ignoring the aspect of motivation suggesting education should inspire and stimulate student curiosity.[12,13] Lineberry puts forward the concept of "Assessment-affecting-learning", which considers student motivation, and recommends using this as the primary

way of encouraging learning and performance in education.[9] But, this approach does not differentiate between driving student learning through controlled and autonomous motivation, and the harmful effect of driving controlled motivation on autonomous motivation.[12,13] Self-determination Theory (SDT) of motivation – which classifies motivation as autonomous (out of genuine interest and /or personally endorsed importance) and controlled (out of internal or external pressure or contingent to rewards or sanctions) – can provide guidance on how assessments should be conceptualized to foster autonomous rather than controlled motivation.[12,13] Autonomous motivation as compared to controlled motivation is associated with deep learning, better academic performance, higher creativity and psychological well-being.[12-14] Stimulating autonomous motivation is contingent on the satisfaction of three basic psychological needs: autonomy (sense of choice in learning), competence (sense of capability for learning), and relatedness (sense of belonging to the peer group). In contrast controlled motivation is stimulated on frustration of these needs .[12-14]

SDT posits that high-stakes assessments have deleterious effects on students' autonomous motivation for learning and can corrupt educational practices.[4] Even after educators distinguish between formative and summative assessments in HPE, students perceive formative moments as summative ones. They try to control their grades by choosing their best performance moments for their formative assessments.[15] This may be rooted in the notion that assessments often are deterministic of future educational opportunities. This can, therefore, happen even in sophisticated assessment systems/programs, like programmatic assessment, which is primarily designed as an *assessment-for-learning* opportunity.[16] In medical schools who have adapted programmatic assessment, students have been found to treat formative assessments like summative assessments when they do not feel a sense of control over the assessment outcome.[17] This is especially true for knowledge-based assessments with structured answers, when there is a lack of a trusting relationship with teachers and when assessments cannot be used for improving performance.[17] Moreover, teachers may implement assessments differently than intended by curriculum developers; e.g. different clinical supervisors have been found to apply standards for scoring competencies differently.[3,18] This is an additional factor that can widen the gap between the intention and implementation of

assessments. Thus, even a well-intentioned assessment system/program, may work against student motivation, if implemented incorrectly .[19]

Therefore, this review aimed to examine the effect of assessments on motivation and the outcomes of these effects, as reported in the HPE literature, thereby attempting to provide a scientific grounding for organizing assessments to stimulate student curiosity and autonomous motivation which, in turn, should ultimately foster learning and long-term clinical performance. The research questions guiding this review were:

How do assessments affect student motivation for learning in HPE? What outcomes does this lead to in which contexts?

Methods

We conducted the search in a systematic manner and used realist synthesis method for data analysis. We used the cited realist reviews published in HPE along with methodology articles to guide our methods.[20-24]

Search strategy – The search strategy was developed iteratively with an information specialist, and was sharpened and limited to the 2010-2020 period.[21] A comprehensive search was performed in the bibliographic databases PubMed, Embase.com APA PsycInfo (via Ebsco), Eric (via Ebsco), Cinahl (via Ebsco) and the Web of Science Core Collection from January 1st 2010 to October 29th 2020, without language restrictions. The following terms were used in 3 iteratively developed search strings (including synonyms and closely related words) as index terms or free-text words: “assessments” AND “motivation” AND “health professions education/students” (See Supplemental Digital Appendix 1 for the full search strategy). Duplicate articles were excluded. A snowball search was conducted on the references of all included articles to identify more relevant articles.

Inclusion criteria - a) investigating the effect of assessments (all types) on student motivation in HPE, b) empirical papers and literature reviews, c) quantitative, qualitative and mixed methods, and d) all languages.

Exclusion criteria - a) Non-HPE population, b) Not on assessments and motivation, c) Non-empirical publications - opinions, perspectives, letters, editorials, commentaries, dissertations, conference abstracts, and d) motivation was not an outcome measure or was not measured/evaluated.

Rigor and relevance – The rigor and relevance of included articles were evaluated as specified by the RAMSES standards of publication for a realist synthesis.[20,21] Rigor evaluated the credibility and trustworthiness of the method used to generate the results. Relevance evaluated the importance of the article in answering our research questions.[20]

Rationale for choosing realist synthesis – We wanted to study the intended and unintended consequences of a complex topic - the effects of assessments on motivation and hence learning outcomes . Because realist synthesis explores "which mechanisms lead to what outcomes in which contexts" (Context-Mechanism-Outcome), it seemed to be the most suitable method for our analysis. [20,21] That said, we did not follow the realist review method described by Carrieri et al. because our intention was not to find an immediately practically applicable intervention (by involving important stakeholders at all stages).[25] Our intention was, instead, to study the context-mechanism-outcomes of factors in assessments that influence motivation and hence produce learning and psychological well-being outcomes. Thus, the realist method followed other authors, who also studied factors influencing certain variables in HPE, suited our objectives better.[21-24]

Data extraction – RAK and CO or SS read all titles and abstracts to make inclusion-exclusion decisions, and extracted data on "context-mechanism-outcome" (See Supplemental Digital Appendix 2).

Data analysis –

RAK conducted all steps of the analysis. CO and SS checked each step independently. AA checked the coded data of 6 randomly chosen articles. Firstly, RAK used MaxQDA software program to conduct a content analysis of the data using sensitizing concepts from SDT to identify stimulation of autonomous and controlled motivation.

Secondly, she classified the codes of content analysis into context, mechanism and outcome. The conditions which led to effects on autonomous or controlled motivation constituted the context. How autonomous or controlled

motivation was stimulated constituted the mechanism. Outcomes comprised learning and psychological well-being, among others..

Thirdly, RAK extracted context-mechanism-outcome configurations to propose overarching program theories on how factors stimulate autonomous or controlled motivation. The realist program theories proposed in the article were finalized through consensus in the whole research team.

Results

On applying the inclusion-exclusion criteria, 24 out of 15,291 articles were included (See Figure 1). The rigor and relevance of each article was reported (See Supplemental Digital Appendix 3).

Ten studies had been conducted on medical students or residents, 4 on nursing students, 1 on medical and nursing students, 2 on nutrition students, and 1 each on veterinary medicine, physiotherapy, health sciences, oral health, physician assistant, pharmacy and dental students. Eleven studies had been conducted in Europe, 4 in Australia, 3 each in Asia and USA, 2 in Canada and 1 in the Middle East. As we used a realist synthesis approach, we focus on the features of the different assessments in this review. The specific assessments are, therefore, not relevant, but have been included in as Supplemental Digital Appendix 4.

In our analysis we could identify which assessment features stimulated autonomous versus controlled motivation. We report the context, mechanism and outcomes for each of these two broad categories below (See Table 1 for an overview of the main findings):

Assessment features/principles stimulating controlled motivation

The outcomes of assessments stimulating controlled motivation included higher effort driven by external pressure,[26-30] surface and strategic learning (learning only what is likely to be assessed) ,[26-28,31-34] which happened at the expense of learning out of interest or for future practice, undermining of student engagement and learning,[34]

poorer psychological well-being (feeling of pressure to study, anxiety, frustration and stress),[26,27,33] creation of a 'performance culture' at the expense of learning and development,[33] and undermining of the learning potential of an assessment program.[33] Failure in assessments drove away inspiration and motivation,[35] reduced students' sense of self-efficacy and colored their reflections of workplace based learning.[33]

Stimulation of controlled motivation happened in different contexts. The mere presence of assessments (formative[31,32] or summative[26-28,36,37]) was an external motivator for learning or for preparing for the assessments. Further contexts were the presence of grades,[34,38] high stakes assessments,[26] like Licensure exams,[36]strong focus on factual knowledge,[27] the competitive element,[26,31] pressure related to performance,[27] time,[26-28,35] emotions,[27] external reward,[27] and internal reasons, i.e. the desire to fulfill supervisor's expectations,[36] fear of disclosing lack of knowledge or failure, [26,27] conflict between preparing for the clinical profession and to pass the final examinations,[26,38] and consequences being coupled with assessments.[39] Timing of assessment seemed to play a role as motivation seemed to dip towards the end of the assessment.[40]

Assessments stimulated controlled motivation through several mechanisms. *External pressure exerted to study in orderto* learn material that was important for practice,[28] and also to study and practice with a goal to pass assessments stimulated controlled motivation.[27,31] Examinations provided a framework and forced students to acquire and prioritize the required knowledge.[28] External pressure and reward functioned as triggers for controlled motivation.[27,29,31,38,41,42] Assessments triggered controlled motivation which further made students study and practice for assessments.[27] Some formative assessments types stimulated students' controlled motivation and hence achievement.[31] Grading of assessments motivated students to submit high quality work.[43] Questions asked in a formative assessment, were perceived it to be important for final examinations, and therefore garnered more study effort.[31] Although assessments made students felt pressurized to study more , they also stimulated liking for the subject.[27] The existence of external regulating bodies and regulations stimulated lifelong learning.[36] Focus on entrustment as the assessment outcome seemed to enhance learner perceptions of being judged continuously and of assessments being perceived as summative rather than formative.[33] This effect can undermine the learning potential of an assessment program which is based on trust and psychological safety.[33] Because of its focus on autonomy

and unsupervised practice, entrustment language can lead to a 'performance culture' at the expense of learning and development.[33] Assessments also led to anxiety and frustration.[26,27,40,41]*Assessments also stimulated controlled motivation through creating competition or out of fear for poor performance.* Peer assessment led comparison of performance with peers,[42] fear of receiving low marks (if they did not contribute)[29] and motivation to learn. Wanting to know what their peers thought of their work, prompted them to reread their work in light of the peer feedback.[42] Students studied harder for the reward of extra credits,[27] the fear of getting lower grades or failure[26,29] and poor performance.[30] Competition motivated students to prepare for assessments.[26,31] Students studied hard in collaborative testing to avoid disappointing fellow students. This also strengthened their confidence and made them excel.[44] *Assessments stimulated controlled motivation resulting in encouragement of surface and strategic learning.* Focus on factual knowledge in assessments and controlled motivation led to surface learning.[27] Students who learned by rote memorization without understanding and regurgitated the material during the exams tended to fare better on certain assessments.[26] When traditional grading was employed for assessments, students valued the assessment in the hierarchy of the weightage carried by it in the total picture. This influenced their engagement with the topic and the learning objectives.[34] Choosing what to study was based on previous exam questions rather than on the knowledge essential for practice.[28] Students were inclined to pick easier assessment activities in order to fare better.[26] High stakes assessments encouraged a surface learning approach, while other assessment types encouraged a deep learning approach owing to the lower stakes.[26] Making certain assessment types (like Entrustable Professional Activities - EPA) high stakes may give rise to strategic behavior of students.[33]*Individual variation in perceptions about assessments gave rise to different behaviors.* Some students perceived generating ideas as a competitive process, while others perceived it as an opportunity for reflection.[40]

Assessment features/principles stimulating autonomous motivation

The outcomes of assessments that stimulated autonomous motivation were higher creativity[34], skill development,[44] fostering of self-regulation, time management skills, mastery learning and self-directed corrective actions,[30,34,41] interest in the topic,[27] better learning approach,[26,43,45,46] higher effort and better connection with the

class material,[34,47] increased quality of work,[29] competence, confidence and performance,[26,40] goal setting and engagement in learning,[33,44] better error identification and higher effort to practice clinical skills,[48] increased motivation to spend more time with patients,[38] skills and professionalism development,[45] and encouragement for work improvement.[42] Lower effort was found in formative compared with summative assessments in spite of formative assessments stimulating autonomous motivation.[38] Reflection on assessments enhanced student motivation, learning and well-being.[33]

Stimulation of autonomous motivation happened in the context of the assessment being fun, interesting, and challenging,[26,27,36,40,45] innovative, mastery-based and conducted longitudinally,[33,34] introduced in the study in a seamless manner,[46] and embedded with facets of student autonomy, e.g. students graded themselves on mastery criteria pre-determined by the teacher,[34] use of pass/fail grading only,[34] provision of multiple attempts to pass,[26] choices of assignments,[34] choice in learning their topics of interest,[40,45] and choice in place and time of assessment.[39] *Stimulation of autonomous motivation also happened* if the assessment was poised to directly build *competence*, and embedded active learning assessment, with timely, external and multiple sources of feedback.[28,31,33,40,41,44,47,49] Assessments which included setting individual[40] or shared goals,[41] stimulated learners to reflect,[33] were jointly constructed between learners and teachers,[45] had authenticity with clinical practice,[26,28,35,37,38,49] encouraged sharing or shared/group learning and responsibility,[27,39,41,44] and held personal value for students also stimulated autonomous motivation.[27,31,45] Certain student characteristics also form the context for autonomous motivation stimulation, such as baseline motivational levels,[39] having the motivation for personal achievement or satisfaction or having a 'love of learning' and passion for the topic.[26,40] Variations in the effects of assessments are seen depending on the type of assessment, i.e. testing fact recall or deep thinking.[26]

Assessments stimulated autonomous motivation through several mechanisms. A fun and challenging assessment, on its own, triggered autonomous motivation and provided a holistic picture which would be difficult for a student to do on his own.[27] The students were more driven by personal motivation and changed their learning approach.[45] Students worked to a schedule because of their autonomous motivation.[32] Students would rise to the assessment occasion as they just wanted to be good doctors.[38] Formative assessments motivated students by making them

aware of what they knew and what to study.[27] Particular assessment types stimulated autonomous motivation by providing instant feedback through rubrics and shifting the focus to mastery learning.[34] *Autonomy in learning also stimulated autonomous motivation.* Use of portfolio made students approach learning through a more personal motivation.[45] Choice (e.g. of topics to study/add in the portfolio) in assessment removed boundaries, allowed for personal exploration and increased students' appreciation of the topics of study.[45] As students progressed through the planning process and became more independent, their motivation levels increased.[45] Autonomy in learning made students independently develop their own skills and increased their motivation.[40] Being able to watch their own performance motivated students to practice their skills.[48] Certain assessment types motivated students to study more, focus on important concepts, and reflect on their learning.[31] Some assessment types increased motivation by giving immediate feedback.[41] Knowing and understanding what was expected in the assessment helped students to improve their assignments.[42] *Getting feedback stimulated autonomous motivation in several ways*. Face to face feedback improved students' competence and confidence.[26] Getting feedback from patients and colleagues motivated students.[49] Getting to know the correct answer immediately after the assessment, stimulated students to focus more on the questions and stimulated their motivation.[31] Assessment followed by explanatory feedback enhanced the learning process and sustained student motivation.[46] Error detection helped students to identify their knowledge gaps.[47] Collaborative testing helped in closing a performance gap through constant peer evaluation and feedback.[44] By providing a better picture on their development, reflection on assessments improved students' motivation, learning and well-being.[33] *Active learning assessments stimulated autonomous motivation.* They led to increase in the skills and confidence and increased motivation after the completion of each learning journey stage.[40,47] Active learning assessments not only supported knowledge building but also engaged students cognitively and emotionally.[42,45] They generated enquiry by providing students the opportunity to formulate questions.[45]*Assessments having authenticity with practice motivated students intrinsically* as they provided a sense of wholeness, fostered clinical skills and professionalism, captured students' interest and encouraged team work.[27,38,41,44,45] Choices in portfolio allowed students to include their personal experiences into their study and significantly motivated them.[45] Professional responsibility as physicians motivated students for lifelong learning.[36]

Peer/group/shared learning helped students to identify gaps in their knowledge, created constructive friction and moved them into Vygotsky's Zone of Proximal Development (i.e. the zone in which students have enough challenge and are motivated to learn new things), enhancing their learning.[41] Peer assessment functioned as an extra motivational strategy for individual students to contribute to the group in a meaningful way and provided teamwork experience.[29,44] Engagement in peer learning led to interaction, resulting in engagement and motivation.[44] The peer assessment process empowered students, motivated them, and increased their confidence and engagement in learning.[42] Group assessment helped students to get a helicopter view of topics, which would have been hard to achieve on their own.[27] Learning in a group with a shared interdependent goal made students feel personal responsibility for the group work and bring more knowledge and effort in their contribution.[41] Some students perceived generating ideas as a competitive process, while others perceived it as an opportunity to reflect on their ideas and making comparisons.[41] Failure led students to study harder.[33] Not passing their entrustment assessment made students feel frustrated.[33]

Emergent program theories from the context-mechanism-outcome combinations

Figure 2 depicts the emergent realist program theory that explains how assessments enhance controlled motivation and lead to negative learning and psychological well-being outcomes. Negative psychological well-being was an outcome reported only in qualitative data and was characterized by negative emotions such as anxiety, stress and frustration.[26,27,33] For an understanding on how assessments enhance controlled motivation, we found contextual factors at an assessment system/program level as well as at an individual student level that work through taking away autonomy and creating negative perception of competence. We found creation of a performance culture and undermining of the learning potential at an assessment system/program level (See Figure 2).

Figure 3 depicts the emergent realist program theory that explains how assessments enhance autonomous motivation and lead to positive learning outcomes. For insights into how assessments enhance autonomous motivation, we found that contextual factors in the assessment features, i.e. in the assessment content and format,

as well as at an individual student level, work through satisfying the basic psychological needs of autonomy, competence and relatedness, and creating value for the activity to produce positive learning outcomes at an individual student level (See Figure 3). We did not find psychological well-being outcomes for assessments that stimulate autonomous motivation.

Discussion

In this review, we add to the HPE literature by describing context-mechanism-outcome configurations for how assessments influence motivation leading to individual learning and psychological well-being outcomes as well as systems/program-level outcomes. In general, we found that assessments and their features used in HPE predominantly stimulate students' controlled rather than autonomous motivation. We found that assessments, at the system/program as well as individual student levels, enhance controlled motivation by frustrating the SDT-based psychological needs of autonomy and competence, leading to negative consequences at the systems/program as well as individual student levels. In contrast, we found the assessments characteristics as well as individual student characterstics enhance autonomous motivation through satisfaction of the SDT-based basic psychological needs of autonomy, competence and relatedness, along with creation of 'value' to produce outcomes only at an individual student level. We did not find any psychological well-being outcomes or systems/program level outcomes for enhancing autonomous motivation in the current literature.

In Table 2, based on our findings, we have provided recommendations for educators on converting assessments stimulating controlled motivation into ones that can stimulate autonomous motivation. This is important because assessments that stimulate controlled motivation not only produce negative psychological well-being, but also have a long-term deleterious effect on autonomous motivation for learning.[4]

In current HPE practice, the two major assessment concepts being implemented widely are programmatic assessment and EPAs.[50] We did not find any research on the effect of programmatic assessment on student

motivation. Such research would add to the literature, especially because programmatic assessment has some of the features identified in this review that seem to stimulate autonomous motivation.[51] In terms of EPAs, we found one study suggesting that EPA-based assessments mainly stimulated controlled motivation.[33] This finding was somewhat surprising, given that EPA-based assessments have some of the features identified in this review as stimulating autonomous motivation.[52] We hypothesize that this disconnect may be due to the gap between the design of individual EPA assessments and how they are embedded in an assessment program or an overall assessment culture, but this needs further investigation.

The gap between the intended and implemented assessment, which has been highlighted throughout the results of this review, aligns with earlier published literature.[4] Tensions between different assessment practices, the implementers' beliefs about assessments and the requirement to uphold and guarantee quality of graduating trainees may account for this gap.[53] This issue can be addressed by a multi-pronged strategy including: considering student motivation when designing assessments, changing the culture within which assessments are embedded, training faculty to implement assessments *as they are intended* [3], and building a shared understanding between teachers and students on the goals of assessment.[54]

Implications for practice - The features of assessments that we found can stimulate autonomous motivation can be used by educators for redesigning current assessments or developing innovative assessments. We recommend developing assessments that more authentically prepare students for clinical practice, such as having more clinical reasoning exercises where students describe patient problems and the underlying mechanisms, or diagnostic justification exercises where students suggest a differential and rationale, over the standard MCQs. We found a disturbing trend which suggests that students strategically learned what was expected in the assessments at the expense of what is needed in health professions practice. This is more alarming in light of SDT's claim that stimulation of controlled motivation through incentives (grades and qualifications) has a deleterious effect on autonomous motivation.[4,55] Only if we pay attention to stimulating autonomous motivation for learning through assessments, will we be able to "light the fire of learning" instead of just "filling the bucket" for HPE students.[13]

Implications for further research – The results of this review provide a scientific basis for developing a research program on: designing innovative assessments stimulating autonomous motivation and investigating how they work through design-based research, and the effects of programmatic and EPA-based assessments on student motivation.

Limitations – This review was limited to the HPE literature. A broader review of effects of assessments on motivation in higher education may be beneficial. Although we evaluated the rigor and relevance of each study in the review, we did not use this evaluation to exclude articles. This approach, however, fits well with the realist synthesis method, allowing us to include all ideas that contribute to innovative assessments methods.

Conclusion

Assessments in HPE, included in this review, seem to stimulate mainly controlled instead of autonomous motivation of students, with many resulting in difficult consequences like negative psychological well-being. Our findings indicate that students strategically learn what is expected in the assessments at the expense of what is needed in the practice. This approach leads to stress and negative psychological well-being. Therefore, health professions educators urgently need to rethink their assessment philosophy and practices and introduce assessments that stimulate curiosity and genuine interest in the content and for professional practice.

References

1. Wormwald BW, Schoeman S, Somasunderam A. Assessment drives learning: An unavoidable truth? Anat Sci Educ. 2009;2:199-204.
2. Sluijsmans D, Dochy F, Moerkerke G. Creating a learning environment by using self-, peer- and co-assessment. Learn Environ Res. 1998;1(3):293-313.
3. Schellekens LH, Bok HG, de Jong LH, van der Schaaf MF, Kremer WD, van der Vleuten CPM. A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). Stud Educ Eval. 2021;71:101094.
4. Ryan RM, Weinstein N. Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. Theory Res Educ. 2009;7(2):224-233.
5. Lyndon MP, Strom JM, Alyami HM. Yu TC, Wilson NC, Singh PP, Lemanu DP, Yielder J, Hill AG. The relationship between academic assessment and psychological distress among medical students: a systematic review. *Perspect Med Educ.* 2014;**3**:405–418.
6. Warr P. 2012. How to Think About and Measure Psychological Well-being. In: Research Methods in Occupational Health Psychology. Eds: Wang M, Sinclair RR. Tetrick LE. New York: Psychology Press/Routledge.
7. Huppert FA. Psychological Well-being: Evidence Regarding its Causes and Consequences. Applied Psychology: Health and Well-being 2009;1(2):137–164.
8. Schunk DH. Learning theories: An educational perspective. (6th. Ed.). Boston, MA: Pearson; 2012, pg. 3.
9. Lineberry M. Assessment affecting learning. In: Assessment in Health Professions Education. Eds. Yudkowsky R, Park YS and Downing SM. 2020, 2nd edition, Routledge, New York, ©Taylor and Francis. Pgs. 257-268.
10. Scott IM. Beyond 'driving': The relationship between assessment, performance and learning. Med Educ. 2020;54:54–59.
11. Norcini J, Brownell Anderson M, Bollela V, Burch V, Costa MJ, Duvivier R, Hays R, Mackay MFP, Roberts T, Swanson D. 2018 Consensus framework for good assessment. Med Teach. 2018;40(11):1102-1109.

12. Deci EL, Ryan RM. Self-Determination Theory: A macrotheory of human motivation, development, and health. Can Psychol. 2008;49(3):182-185.
13. Kusurkar RA, Ten Cate ThJ. AM Last Page: Education is not filling a bucket, but lighting a fire: Self-Determination Theory and motivation in medical students. Acad Med. 2013;88(6):904.
14. Kusurkar RA. Self-determination Theory in Health Professions Education Research and Practice. In: The Oxford Handbook of Self-determination Theory. Ed. Ryan RM. ©Oxford University Press, NY. (In Press)
15. Daelmans H, Mak-van der Vossen M, Croiset G, Kusurkar RA. Faculty-perceived difficulties and needs in workplace based assessments: A report. Int J Med Educ. 2016;7:19-24.
16. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Bartman LKJ, Van Tartwijk J. A model for programmatic assessment fit for purpose. Med Teach. 2012;34:205–214.
17. Schut S, Driessen E, Van Tartwijk J, Van der Vleuten CPM, Heeneman S. Stakes in the eye of the beholder A international study of learners perceptions within programmatic assessments. Med Educ. 2018;52:654-663.
18. Westein MPD, Koster AS, Daelmans HEM, Collares C, Bouvy ML, Kusurkar RA. Validity, reliability and impact on decision making of summative performance evaluations: a mixed methods study in postgraduate pharmacy education. Curr Pharm Teach Learn. 2022;14(6):701-711.
19. Kusurkar RA, Croiset G. Autonomy support for autonomous motivation in medical education. Med Educ Online. 2015;20(1):27951.
20. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. BMC Med Educ. 2013;11:21.
21. Barry M, de Groot E, Baggen Y, Smalbrugge M, Moolenaar N, Bartelink MEL, Damoiseaux R, Scherpbier N, Kluijtmans M. Understanding the broker role of clinician–scientists: A realist review on how they link research and practice. Acad Med. 2019;94(10):1589-98.
22. Stammen LA, Stalmeijer RE, Paternotte E, Oudkerk Pool A, Driessen EW, Scheele F, Stassen LPS. Training physicians to provide high-value, cost-conscious care: A systematic review. JAMA 2015;314(22):2384-2400.

23. Paternotte E, van Dulmen S, van der Lee N, Scherpbier AJAA, Scheele F. Factors influencing intercultural doctor–patient communication: A realist review. Pat Educ Couns. 2015;98(4):420-445.
24. Kors J, Paternotte E, Martin L, Verhoeven C, Schoonmade L, Peerdeman SM, Kusurkar RA. Factors influencing Autonomy Supportive Consultation: A Realist Review. Pat Educ Couns. 2020;103(10):2069-2077.
25. Carrieri D, Mattick K, Pearson M, Papoutsi C, Briscoe S, Wong G, Jackson M. Optimising strategies to address mental ill-health in doctors and medical students: 'Care Under Pressure' realist review and implementation guidance. BMC Med. 2020;18:76.
26. Cobb KA, Brown G, Jaarsma ADC, Hammond RA. The educational impact of assessment: A comparison of DOPS and MCQs. Med Teach. 2013;35(11):e1598-e1607.
27. Weurlander M, Söderberg M, Scheja M, Hult H, Wernerson A. Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. Assess Eval High Educ. 2012;37(6):747-760.
28. Weller JM, Henning M, Butler R, Thopmson A. The impact of the Australian and New Zealand College of Anaesthetists' specialist examinations on trainee learning and wellbeing: a qualitative study. Anaesth Intens Care 2014;42:736–744.
29. Elliott N, Higgins A. Self and peer assessment – does it make a difference to student group work? Nur Educ Pract. 2005;5:40–48.
30. Vinall R, Kreys E. Use of end-of-class quizzes to promote pharmacy student self-reflection, motivate students to improve study habits, and to improve performance on summative examinations. Pharm. 2020;8(3):167.
31. Ismail MAA, Ahmad A, Mohammad JAM, Fakri NMRM, Nor MZM, Pa MNM. Using Kahoot! as a formative assessment tool in medical education: a phenomenological study. BMC Med Educ. 2019;19:230.
32. Ahmad HF, Asif M. Medical students' learning habits: A mixed method study during clinical rotation in general surgery. J Pak Med Asso. 2018;68:600-606.
33. Martin L, Sibbald M, Vegas DB, Russell, Govaerts M. The impact of entrustment assessments on feedback and learning: Trainee perspectives. Med Educ. 2020;54(3):328–336.

34. Pope L, Parker HB, Ultsch S. Assessment of Specifications Grading in an Undergraduate Dietetics Course. J Nutr Educ Behav. 2020;52:439−446.
35. Bengtsson M, Ohlsson B. The nursing and medical students motivation to attain knowledge. Nur Educ Tod. 2010;30:150–156.
36. Sockalingam S, Soklaridis S, Yufe S, Rawkins S, Harris I, Tekian A, Silver I, Wiljer D. Incorporating lifelong learning from residency to practice: A qualitative study exploring psychiatry learners' needs and motivations. J Contin Educ Health Prof. 2017;37(2):90-97.
37. Alkhamees M, Al-Zahrani MA, Almutairi S, Alkanhal H, Almuhaideb M, Abumelha SM. Study habits of urology residents in Saudi Arabia: Identifying defects and areas for curricular development – A trainee-based survey. Adv Med Educ Pract. 2020;11:707–715.
38. Tey C, Chiavaroli N, Ryan A. Perceived educational impact of the medical student long case: a qualitative study. BMC Med Educ. 2020;20:257.
39. Schüttpelz-Brauns K, Hecht M, Hardt K, Karay Y, Zupanic M, Kämmer JE. Institutional strategies related to test-taking behavior in low stakes assessment. Adv Health Sci Educ. 2020;25:321-335.
40. Paxman JR, Nield K, Hall AC. Motivation, confidence, and control; Unraveling active learning for nutrition and food undergraduates. J Food Sci Educ. 2011;10:45-53.
41. Asghar A. Reciprocal peer coaching and its use as a formative assessment strategy for first-year students. Assess Eval High Educ. 2010;35(4):403-417.
42. Casey D, Burke E, Houghton C, Mee L, Smith R, van der Putten D, Bradley H, Folan M. Use of peer assessment as a student engagement strategy in nurse education. Nurs Health Sci. 2011;13:514–520.
43. Harris JR. Peer assessment in large undergraduate classes: an evaluation of a procedure for marking laboratory reports and a review of related practices. Adv Physiol Educ. 2011;35:178–187.
44. Saunders A, Say R, Visentin D, McCann D. Evaluation of a collaborative testing approach to objective structured clinical examination (OSCE) in undergraduate nurse education: A survey study. Nur Educ. Pract. 2019;35:111-116.

45. Leadbeatter D, Gao J. Engaging oral health students in learning basic science through assessment that weaves in personal experience. J Dent Educ. 2018;82(4):388-398.
46. Makransky G, Mayer R, Nøremølle A, Cordoba AL, Wandall J, Bonde M. Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. Educ Tech Res Dev. 2020;68:293–317.
47. Yoneda M, Yamada K, Izumi T, Matsuzaki E, Maruta M, Hatakeyama J, Morita H, Tsuzuki T, Anan H, Hirofuji T. Development of an error-detection examination for conservative dentistry education. Clin Exp Dent Res. 2020;6:69–74.
48. Sadowski CK, Dickerson L, Taylor LF. Student self-assessments and peer assessments of video recordings of physical examination practicums in a physical diagnosis course. J Phys Assis Educ. 2020;31(2):85-90.
49. González-Gil MT, Parro-Moreno AI, Oter-Quintana C, González-Blázquez C, Martínez-Marcos M, Casillas-Santana M, Arlandis-Casanova A, Canalejas-Pérez C. 360-Degree evaluation: Towards a comprehensive, integrated assessment of performance on clinical placement in nursing degrees: A descriptive observational study. Nurs Educ Tod. 2020;95:104594.
50. Holmboe ES, Yamazaki K, Hamstra SJ. The evolution of assessment: Thinking longitudinally and developmentally. Acad Med. 2020;95(11S):S7-S9.
51. Van der Vleuten CPM, Heeneman S, Schut S. Programmatic Assessment: An Avenue to a Different Assessment Culture. In: Assessment in Health Professions Education. Eds. Yudkowsky R, Park YS, Downing SM. 2nd edition, 2019, ©Routledge, New York, USA. Pgs. 245-256.
52. Ten Cate ThJ, Taylor DR. The recommended description of an entrustable professional activity. AMEE Guide No. 140. Med Teach. 2021;43(10):1106-1114.
53. Govaerts MJB, van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: moving beyond either–or thinking. Med Educ. 2019;53:64–75.

54. Schut S. Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten CPM, Driessen E. Where the rubber meets the road - An integrative review of programmatic assessment in health care professions education. Perspect Med Educ. 2021;10:6–13.
55. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull. 1999;125(6):627-668.

Figure Legends

Figure 1 – HPE – Health Professions Education, n - number

Figure 2 – Legend not required

Figure 3 – Legend not required

Table 1 Overview of the main findings – Context-Mechanism-Outcome for assessment features stimulating autonomous and controlled motivation[1]

Assessment features stimulating controlled motivation		
Context	**Mechanism**	**Outcome**
- The mere presence of a system for assessments - The presence of grades - High stakes assessments - Strong focus on factual knowledge - The competitive element - Performance or time or emotional or external or internal pressure - Fear of failure or disclosing lack of knowledge - Conflict between preparing for clinical profession and to pass final examinations - When consequences are coupled with assessments Timing of assessment – Motivation dipped towards the end of assessment	Through - Forcing to spend time on learning for tests - Study to pass assessments - Providing a framework - Comparison with peers - Fear of failure - Being judged continuously leading to anxiety and frustration - Encouragement of surface or strategic learning Timing of assessment influenced student effort Individual perceptions influenced how students interacted with assessments	- Higher effort driven by external pressure - Surface and strategic learning - Undermining of student engagement and learning - Poorer mental well-being - Feeling of pressure to study, anxiety, frustration and stress - Creation of a 'performance culture' at the expense of learning and development - Undermining of the learning potential of an assessment program - Failure in assessments drove away inspiration and motivation - Reduced students' self-efficacy and colored their reflections of workplace based learning
Assessment features stimulating autonomous motivation		
Context	**Mechanism**	**Outcome**
Assessment was - Interesting, challenging - Innovative, mastery-based and longitudinal - Introduced in a seamless manner as part of instruction - Embedded with facets of student autonomy - Directly building competence - Including setting individual/shared goals - Jointly constructed - Having authenticity with clinical practice - Encouraging shared or group learning and responsibility - Holding personal value for students Certain student characteristics such as baseline motivation/passion for topic	Through - Personal motivation - Autonomy in learning - Feedback - Active learning - Authenticity with practice - Peer or group or shared learning	- Higher creativity - Skill development - Fostering of self-regulation, time management, mastery learning and self-directed corrective actions - Interest in the topic - Better learning approach - Higher effort - Increased quality of work, competence, and confidence - Goal setting and engagement in learning - Better error identification and higher effort for clinical skills - Increased student motivation to spend time with patients - Developing skills and professionalism - Improvement in work

[1]*Context means the circumstances under which a certain effect is observed, Mechanism means the process through which an effect takes place, Outcome means the results of the context and mechanisms.*
A factor mentioned in the context in one article could be mechanism or outcome in another article depending on how it is described.

Table 2 How we can convert assessments to stimulate autonomous instead of controlled motivation by making changes in the assessment features

Assessments features stimulating controlled motivation	**Assessments features stimulating autonomous motivation**
Uninteresting, low challenge	Interesting, challenging
Abstract, theoretical	Authentic, practice-based
Irrelevant to practice	Relevant and transferable to practice
Constructed by teachers	Co-constructed by teachers and students
Individual	Collaborative with shared learning
Implemented per course/ semester	Implemented longitudinally in the study
Do not hold personal value for students	Hold personal value for students
Do not support student autonomy	Support and encourage student autonomy, e.g. by providing choices in the type, format, content and timing of assessment
High stakes	Low stakes
Presence of grades	Pass/Fail system without grades
Strong focus on factual knowledge	Focus on practice-relevant knowledge
Performance or Time or Emotional or External or Internal pressure	Reduce all kinds of pressure (This change needs to happen at the cultural or organizational level)
The competitive element	Remove the competitive element
Fear of disclosing lack of Knowledge or failure	Encouraging a culture of 'Mistakes are good' or 'Failure helps you learn'
Conflict between preparing for clinical profession and to pass final examinations	Designing assessments which test knowledge that is essential to practice

The elements in the left column can be replaced with the elements in the right column in order to stimulate autonomous motivation instead of controlled motivation.

Figure

Figure 1 PRISMA statement

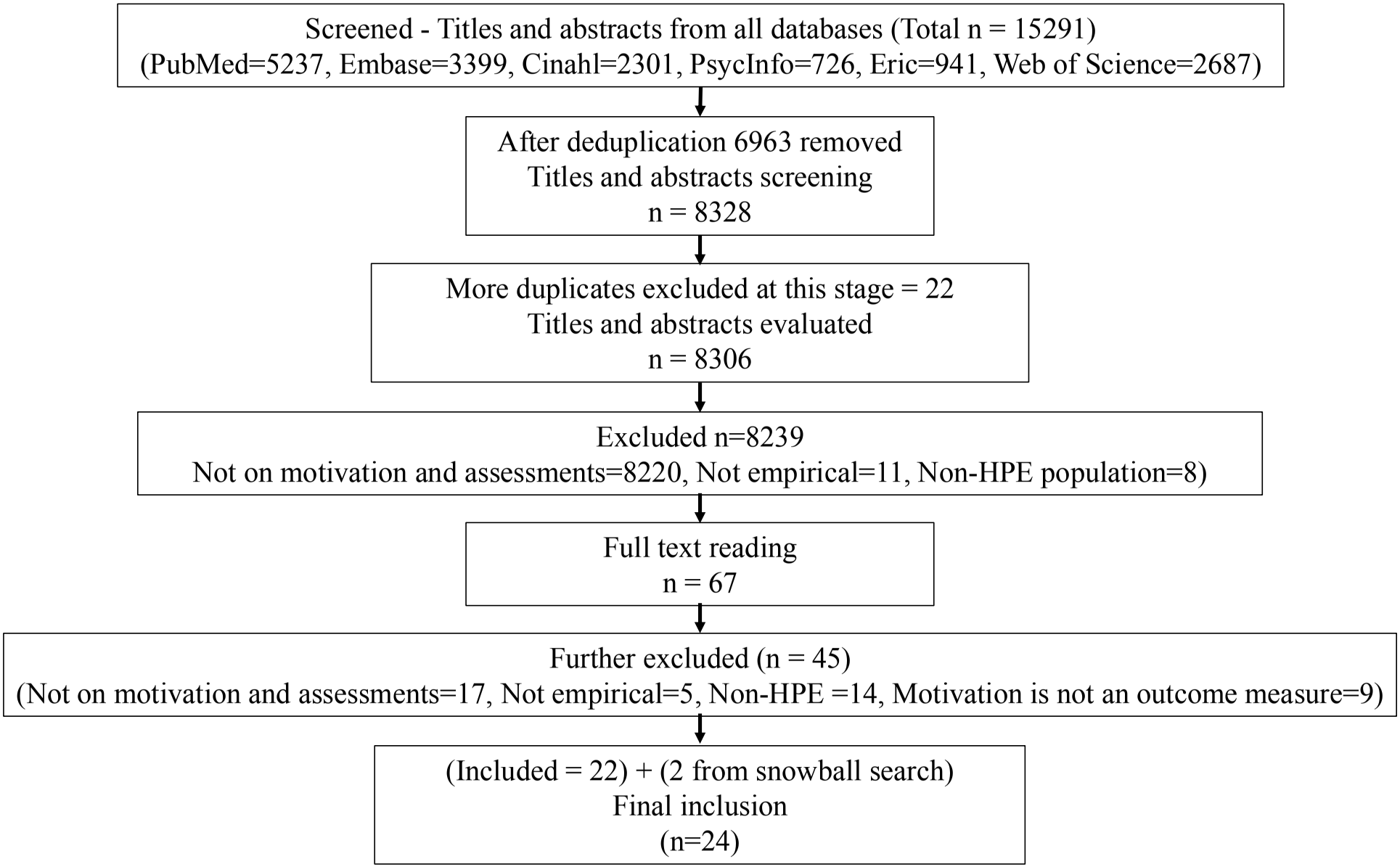
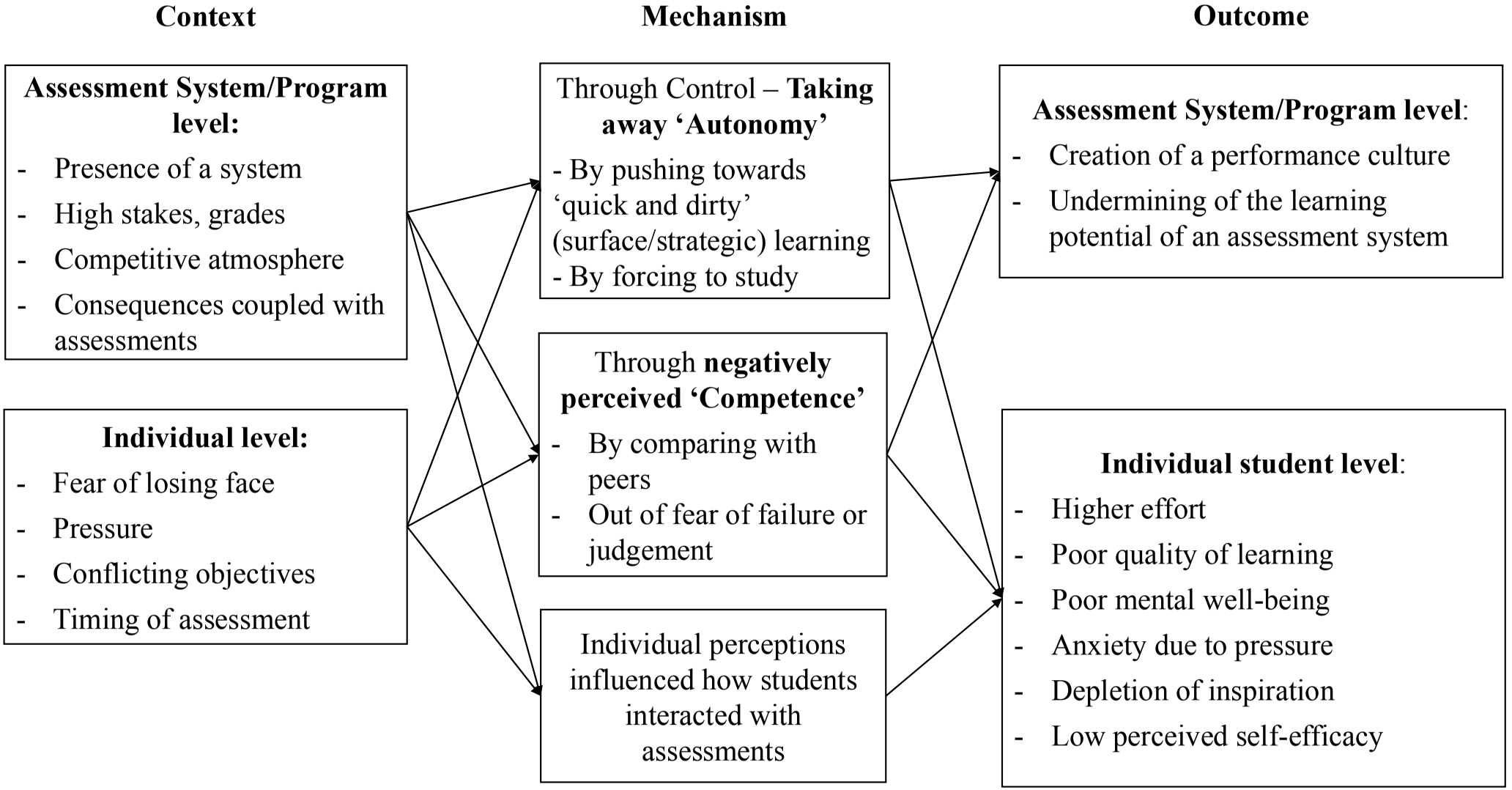


Figure 2 Realist program theory on how assessments enhance controlled motivation and lead to negative learning and psychological well-being outcomes



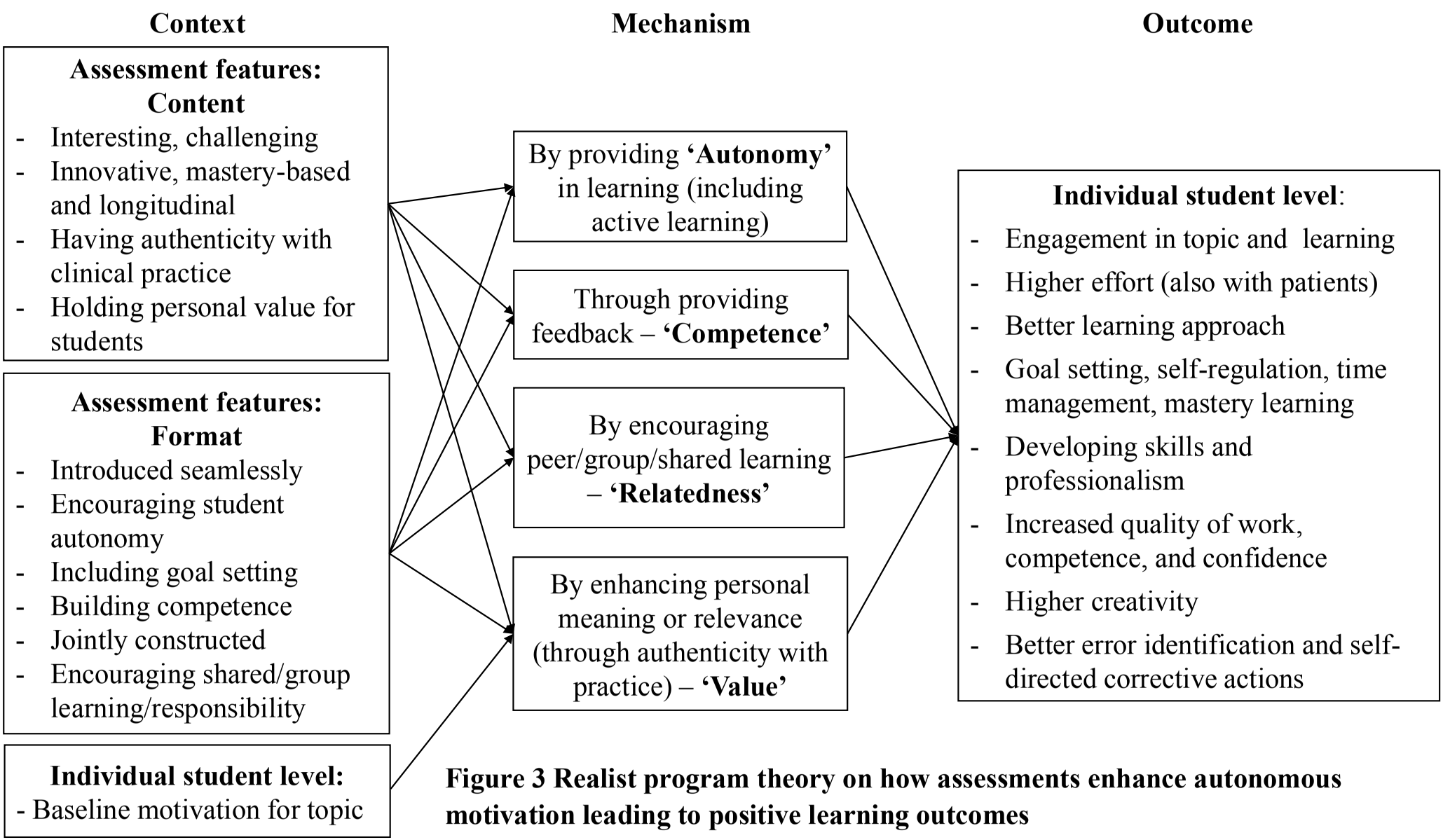


Figure 3 Realist program theory on how assessments enhance autonomous motivation leading to positive learning outcomes