

# There is no “too small” for frequent workplace-based assessment: Differences between large and small residency programs in anesthesia when using a mobile application to assess EPAs

## Abstract

**Background:** A competency-based education approach calls for frequent workplace-based assessments (WBA) of Entrustable Professional Activities (EPAs). While mobile applications increase the efficiency, it is not known how many assessments are required for reliable ratings and whether the concept can be implemented in all sizes of residency programs.

**Methods:** Over 5 months, a mobile app was used to assess 10 different EPAs in daily clinical routine in Swiss anesthesia departments. The data from large residency programs was compared to those from smaller ones. We applied generalizability theory and decision studies to estimate the minimum number of assessments needed for reliable assessments.

**Results:** From 28 residency programs, we included 3936 assessments by 306 supervisors for 295 residents. The median number of assessments per trainee was 8, with a median of 4 different EPAs assessed by 3 different supervisors. We found no statistically significant differences between large and small programs in the number of assessments per trainee, per supervisor, per EPA, the agreement between supervisors and trainees, and the number of feedback processes stimulated. The average “level of supervision” (LoS, scale from 1 to 5) recorded in larger programs was 3.2 (SD 0.5) compared to 2.7 (SD 0.4) ( $p < 0.05$ ). To achieve a  $g$ -coefficient  $> 0.7$ , at least a random set of 3 different EPAs needed to be assessed, with each EPA rated at least 4 times by 4 different supervisors, resulting in a total of 12 assessments.

**Conclusion:** Frequent WBAs of EPAs were feasible in large and small residency programs. We found no significant differences in the number of assessments performed. The minimum number of assessments required for a  $g$ -coefficient  $> 0.7$  was attainable in large and small residency programs.

**Keywords:** competency-based education, CBME, WBA, EPA, entrustment, feedback, decision making, residency program,  $g$  theory, reliability, small business, SME, smartphone, mobile application

Tobias Tessmann<sup>1</sup>

Adrian P. Marty<sup>2</sup>

Daniel Stricker<sup>3</sup>

Sören Huwendiek<sup>3</sup>

Jan Breckwoldt<sup>1</sup>

1 University Hospital Zurich, Institute of Anaesthesiology, Zurich, Switzerland

2 University Hospital Balgrist, Department of Anaesthesiology, Intensive Care Medicine and Pain Therapy, Zurich, Switzerland

3 University of Bern, Institute for Medical Education, Bern, Switzerland

## 1. Introduction

### 1.1. Competency-based medical education (CBME)

CBME and its implementation through the concept of Entrustable Professional Activities (EPAs) has increasing uptake in graduate medical education [1], [2]. An EPA can be seen as a frame through which one or more competencies can be observed during concise, specific tasks in clinical practice [3], [4]. The sum of all EPAs in a specialty represents its complete curriculum and thus all requested competences at a certain point of training.

When combined into a portfolio, sets of complementary EPAs can represent important developmental stages [5] in a holistic, longitudinal approach. First published by Ten Cate 2005 [6], the concept of EPAs not only complemented and facilitated the rise of CBME, but also introduced the important aspect of trust and entrustment decisions for future performance. Evidence shows the use of entrustment-supervision scales rather than generic scores to have a greater clinical relevance, better inter-rater reliability and are easier to use [7], [8]. They do not provide a simple school grade, but an answer to the question: in a similar situation, how much supervision will my resident need next time?

To assess EPAs, a broad array of different workplace-based assessments (WBA) is used. Each WBA should not only serve as a tool for evaluation of trainees' current status, but also support their future development through frequent feedback [9]. This concept of “assessment for learning” [10] has been introduced as one of the key components of a programmatic assessment approach. Furthermore, the results of these frequent WBAs should not be used for high-stake decisions like “pass/fail” of a residency program's year. Contrarily, these should be made in so-called competence committees based on data from a variety of sources [11], [12], [13], [14].

To achieve more frequent and brief assessments, mobile applications have been introduced [15], [16], [17], [18], [19], [20], [21], [22] leading to grossly reduced times spent for a single assessment [23]. Still, a considerable number of assessments are necessary for reliable estimates of a trainee's level of competency. Reliability and validity of a tool are closely related to its design and usage in the appropriate context [24], [25], [26], [27], [28], [29]. No matter for which purpose an assessment is carried out and to which decisions its result might contribute, they have to be based on a reliable, transparent and correct set of information.

Usage of entrustment-supervision scales not only involves less assessor workload, but also reduces the necessary number of assessments by around 50% for Mini-CEX [7]. However, the required numbers for other WBA tools are spread over a wide range [24], [29], [30], [31], [32], [33], and the minimum number of assessments for an app-based WBA of EPAs has not been established yet. The minimum number of assessments could be dependent on a variety of factors, such as the number of different assessors, the interaction between supervisors and trainees, or the size of a residency program [4], [8], [25], [34].

## 1.2. Differences between large and small residency programs

Smaller healthcare institutions play a crucial role in healthcare access for everyone on a global scale, esp. in rural and urban under-served areas [35], [36]. Accordingly, high-quality medical education in smaller institutions is a necessity. Besides their structural conditions (number of trainees and supervisors, case-mix and -volume, available resources) [37], smaller institutions differ in respect to the educational culture [38]. There is no *better or worse* situation, but different opportunities:

While one report found large residency programs benefit from their institutions' better infrastructure [39], other authors did not report significant differences [40], or found superior opportunities in smaller institutions for developing clinical skills, better supervisors' attitudes, and a higher sense of self-efficacy in undergraduate medical education [41], [42], [43]. The more specialized care in large institutions is weighed out by a more generalist and holistic approach in smaller ones. In addition, smaller residency programs might offer better chances

for developing non-clinical skills such as resilience and autonomy, while at the same time also being more demanding in the same respect [44]. The more intimate atmosphere, long-lasting mentorships and less fluctuation of staff [45] could benefit the quality of training [46].

## 1.3. Analogy with non-healthcare sector

Outside the healthcare sector, it has also been shown that small- and medium-sized enterprises (SMEs) face their own needs, difficulties and characteristics concerning workplace-based education, resulting in a specific educational culture [47], [48], [49], [50]. It can be characterized on two dimensions: being a more integrated instead of a more formalized training, taking place in a more enabling than constraining learning environment [51], [52]. Acknowledging the different educational culture in SME led to superior training outcomes [53], [54]. As one review puts it: “strategies which match the way a small business learns are more successful than formal training” [55].

## 1.4. Research question

Frequent WBAs of EPAs in residency programs of different sizes have not been studied yet. Gaining insight into the differences and similarities between smaller and larger programs could help to determine whether and how frequent assessments can be applied over the full range of residency programs.

We therefore compared the usage and results of EPAs assessed with a mobile application between large and small residency programs in anaesthesiology and conducted a reliability analysis providing the minimally required number of assessments and assessors to achieve reliable ratings per trainee.

# 2. Methods

## 2.1. Participants and program categories

The education committee of the Swiss Society of Anesthesiology (SGAR-SSAR) invited all Swiss anesthesia departments with residency programs to participate (free of charge). Over 5 months in 2021 all recorded assessments were analysed. According to the national classification, we defined “large residency programs” as departments providing programs of 3.5 or 3 years (categories A1 and A2), and “small residency programs” as those with programs of 1 or 2 years (categories C and B). The large departments host faculty of >50 persons and belong to hospitals which offer highest-level care in a variety of specialties and usually serve as referral hospitals for their respective region. The smaller departments represent the spectrum of primary and secondary healthcare and feature most of the characteristics used for classification above [37]. For completing residency training, a minimal

time of one year at a “small” institution is required in Switzerland.

## 2.2. Mobile application and EPAs

The mobile application “prEPared” (*precisionED Ltd, Wollerau, Switzerland*) allows assessing EPAs close to real-time with minimal interruption of daily workflow [23]. The trainee initiates the process by selecting the EPA performed. Then both trainee and supervisor, independently assess the complexity of the EPA (“simple” vs. “complex”) as well as the recommended LoS. Through comparison of the assessments a feedback conversation is stimulated which is optionally guided by the app using structured feedback instructions. Each assessment datapoint is stored in the trainee’s ePortfolio which the trainee may share with the supervisors [23].

The introduction of the prEPared app was accompanied with information material such as instructional videos about CBME, EPAs and the app.

For the trial period, we suggested ten common EPAs (see table 1, left column), largely resembling the catalogue of first to second year of the national anesthesia training curriculum [56]. Nevertheless, the entire EPA catalogue was available to choose from in the app (so as to provide the trainees and supervisors with the most benefit for individual education). The levels of supervision ranged from 1 (=observe only) to 5 (=supervise others) [57], [58], [59], as shown in table 2.

## 2.3. Statistics and g theory model

As the primary endpoint, we compared numbers of assessments per trainee and per supervisor. Secondary endpoints were differences in single EPAs assessed, the respective ratings (as defined through the supervisor-assessed level of supervision (LoS)), agreement between trainees and supervisors on the LoS, incidence of feedback stimulated by the assessment, and finally, a reliability analysis providing the minimally required number of assessments and assessors to achieve reliable ratings per trainee in a theoretical model (by using g-theory and d-studies).

For the general comparison of supervisors’ ratings of larger and smaller residency programs Wilcoxon U-tests were performed. Bonferroni correction of the alpha error was applied for multiple testing.

G theory is routinely used for the evaluation of assessment tools by trying to quantify and attribute the amounts of variance for individual facets and their interactions [27], [29].

To achieve a balanced dataset for g theory calculation we selected out of all possible trainee/supervisor combinations a sample of 20 trainees, who had completed sets of 3 EPAs with 4 assessments made by 4 different supervisors each. This decision was made a priori for statistical analysis based on the literature findings regarding minimal numbers for similar assessment tools [29], [30], [31], [32], [33]. 16 of the 20 trainees underwent training in

large residency programs, the other 4 in small residency programs, reflecting the proportions of the overall sample. We calculated the impact of the single EPA, the rating (LoS needed according to the supervisor) and the residency program on the variance of the trainees’ assessments. The g theory model was:  $epa \times assessment \times (trainee : residency \text{ program})$ .

The respective variance components were used in D-studies to determine the minimal number of assessments and assessors, respectively.

A g study ( $epa \times assessment \times trainee$ ) and D-studies were also calculated separately for larger and smaller residency programs to compare them against each other and against the overall calculation. To control for and quantify possible differences and interaction effects, a 3-way split-plot ANOVA was performed with category of residency program (large vs. small) as between group factor and EPA (3 levels) as well as supervisor (4 levels) as within subject’s factors. Dependent variable was the supervisors’ LoS. Only the 20 trainees that were eligible for the g study analysis were considered for this analysis.

G theory and d studies were calculated using G\_String A Windows Wrapper for urGENOVA [60]. All other statistical computations were done with SPSS for Windows version 28 (IBM, Armonk, NY, USA).

## 2.4. Data safety and ethics

Due to the strict data privacy policy of precisionED Ltd. and to avoid any concerns and hesitation to participate, no further data (e.g., age, gender, years of training) was gathered. The study was granted exemption by the Ethical Committee of the Canton of Zurich/BASEC-Nr. Req-2019-00242 (March 21, 2019).

## 3. Results

### 3.1. Descriptive summary of data

Between April 1<sup>st</sup> and August 31<sup>st</sup> 2021, 306 supervisors and 295 trainees used the prEPared app to record 3936 assessments. We excluded 93 incomplete assessments (by the trainee *and* the supervisor, 2.3% of the total sample) from the dataset beforehand. In addition, 197 partially incomplete assessments (either by the trainee, or the supervisor, 4.5% of the total sample) were excluded from the respective analyses.

Overall, 28 of the 53 residency programs (52.8%) in Switzerland participated, of which 15 were large programs and 13 were small. Around three quarters of the supervisors and trainees worked in large programs and, accordingly, three quarters of all assessments were recorded there.

Of 722 registered anesthesia residents in Switzerland in 2021 [61], 41% participated in the study. In the participating residency programs, 496 residents were registered, resulting in a 69% participation rate (58% in large, 78% in small residency programs).

**Table 1: Single EPAs with number of assessments and average ratings per category, in descending order of total n**

EPA name	Measurement	Residency program category			Difference (large-small)	p (2-sided U test)
		Overall	Large	Small		
Induction of general anesthesia	n assessments	947	778	169		
	avg. rating	2.98	3.05	2.66	0.39	< .001 *
	Agreement on rating	0.12	0.13	0.04	0.09	.160
Endotracheal intubation	n assessments	563	425	138		
	avg. rating	2.86	2.93	2.64	0.3	< .001 *
	Agreement on rating	0.1	0.13	-0.02	0.15	.025
Intraoperative management	n assessments	438	367	71		
	avg. rating	3.13	3.28	2.36	0.92	< .001 *
	Agreement on rating	0.02	0.03	-0.05	0.07	.202
Spinal anesthesia	n assessments	410	312	98		
	avg. rating	3.06	3.14	2.81	0.33	< .001 *
	Agreement on rating	0.06	0.05	0.08	-0.03	.728
Emergence from general anesthesia	n assessments	333	269	64		
	avg. rating	3.08	3.16	2.71	0.45	< .001 *
	Agreement on rating	0.1	0.12	0.02	0.1	.192
Preoperative assessment/planning/ consent	n assessments	265	216	49		
	avg. rating	3.55	3.66	3.08	0.58	< .001 *
	Agreement on rating	-0.04	-0.02	-0.13	0.11	.247
Arterial line placement (radial)	n assessments	253	215	38		
	avg. rating	3.51	3.62	2.89	0.73	< .001 *
	Agreement on rating	0.07	0.06	0.14	-0.08	.468
Axillary brachial plexus block	n assessments	250	207	43		
	avg. rating	2.85	2.89	2.7	0.19	.457
	Agreement on rating	0.08	0.04	0.24	-0.2	.033
Misc. EPAs	n assessments	166	90	76		
	avg. rating	2.57	2.98	2.09	0.89	< .001 *
	Agreement on rating	0.07	0.09	0.04	0.05	.588
Laryngeal mask insertion	n assessments	156	100	56		
	avg. rating	3.06	3.23	2.75	0.48	.001 *
	Agreement on rating	0.05	0.07	0.02	0.06	.768
Hand-over	n assessments	155	107	48		
	avg. rating	3.59	3.8	3.1	0.7	< .001 *
	Agreement on rating	-0.13	-0.1	-0.21	0.12	.388
Totals	n assessments	3936	3086	850		
	avg. rating	3.07	3.18	2.67	0.51	< .001 *
	Agreement on rating	0.06	0.08	0.01	0.06	.019

**Table 2: Description of the entrustment-supervision scale**

LoS	Description	Description according to [58]
5	Supervise others	May act as supervisor for junior trainees for this EPA.
4	Distant supervision	May practice EPA unsupervised (under distant oversight). Supervisors might be available by phone call etc.
3	Indirect supervision	May practice EPA under indirect (reactive) supervision. Supervisors are not necessarily in the room, but quickly available.
2	Direct supervision	May practice EPA under direct proactive supervision. Supervisors are typically in the same room.
1	Observation only	May be present but may not practice EPA.



**Table 3: Number of assessments, trainees, supervisors and resp. averages per category**

Residency program category	Overall totals				Large				Small			
	<i>n</i>	%	<i>median</i>	<i>IQR</i>	<i>n</i>	%	<i>median</i>	<i>IQR</i>	<i>n</i>	%	<i>median</i>	<i>IQR</i>
Assessments	3936	100	49	112.5	3086	78.4	85	265.5	850	21.6	45	81
...with feedback proc.	2261	57.4	40	87.5	1817	58.9	77	122.5	444	52.2	28	32
Trainees	295	100	4	9	253	85.8	10	24	45	15.3	3	1
Supervisors	306	100	7	8.3	223	72.9	9	18	85	27.8	5	4
Average supervisors per trainee			3				3				3	
Average assessments per trainee			8				8				8	
Average assessments per supervisor			7				8				5	

The median number of assessments per trainee was 8, the median number of different EPAs assessed was 4, by a median of 3 different supervisors. In 2261 (57%) of all assessments, a feedback process was initiated and in 2165 (96%) of those, a learning goal was documented (see table 3).

### 3.2. Comparison of large and small residency programs' results

Overall, we found no relevant differences between large and small programs. In specific, we found no significant difference for the number of assessments, trainees, supervisors, average number of supervisors per trainee, average number of assessments per trainee or per supervisor, or incidence of feedback stimulated (see table 3). We also did not find statistically significant differences regarding the agreement between the supervisor's and trainee's rating on the required LoS and the task complexity. Furthermore, the distribution of single EPAs assessed was similar between the groups (see table 1), as was the average rating of complexity.

As one statistical difference ( $p < 0.01$ ) the average supervisors' LoS ratings, were higher in large programs for almost all EPAs, with an average of 0.5 points more. The other significant difference between the groups was the higher amount of “fixed pairs” in smaller residency programs, cf. to paragraph 3.4.

### 3.3. Details on EPAs and resp. ratings

“Induction of anesthesia” and “tracheal intubation” were the two most frequently assessed EPAs, making up for one third of all assessments. The remaining assessments were evenly distributed between the other EPAs (cf. to table 1).

There were no differences in LoS ratings between single EPAs, with an overall average of 3.11 (SD 0.32), i.e., indirect supervision is required. The trainees' self-assessed LoS for all EPAs was slightly lower than the supervisors' ratings, but no statistically significant difference could be shown. The same holds true for the ratings of complexity,

showing a mean of around one third of all assessed events being classified as complex.

### 3.4. “Fixed pairs” and “enthusiasts”

In both sizes of programs we found fixed pairs (i.e., one trainee collaborating with the same supervisor for 4 or more assessments). Of these 317 pairs, 250 (22.4% of 1116) resp. 67 (38.3% of 175) worked in large resp. small programs. The difference was statistically significant ( $\chi^2(1)=20.604$ ,  $p < 0.001$ ). These pairs accounted for 1843 (46.8% of total) of all assessments: 1381 (44.7%) in large and 462 (52.7%) of them in small programs ( $\chi^2(1)=24.68$ ;  $p < 0.001$ ).

Within the trainees and the supervisors the number of assessments highly varied, with single individuals having recorded more than 200 assessments (2 supervisors and 1 trainee from large and 1 resp. 2 from small programs). 3 of the 15 large and 5 of the 13 small programs returned only a few assessments (<34 assessments).

### 3.5. G theory

The results for the g theory models are shown in table 4, yielding a robust result of  $\phi > 0.7$  with the chosen combination of assessments/EPAs/supervisors while attributing 21% of variance to the facet of interest (the trainee), nested in residency program (as dictated by study design). For the 20 trainees that entered the g theory analysis, the 3-way ANOVA revealed no significant differences in entrustability rating between programs, EPAs or supervisors, nor were there any significant interactions.

All respective D studies are shown graphically in figure 1. The main result can be found in figure 1 (dark red line): theoretically sufficient ( $\phi$  coefficient of 0.7) assessments can be expected with at least a random set of 3 different EPAs out of 10, with each EPA rated at least 4 times by 4 different supervisors, resulting in a total of 12 assessments.

A  $\phi$  coefficient of 0.8, which can be interpreted as extraordinarily robust measurement tool, can be expected when assessing random sets of 6 EPAs with 4 different supervisors (again, totalling 12 assessments).

Table 4: G theory results for 3 EPAs, each assessed 4 times by 4 different supervisors

Residency program category	Generalizability model	Phi	Percentage of variance res. program	Percentage of variance trainee: res. program
All	(trainee: res. program) x epa x assessment	.704	0%	21%
Large programs only	trainee x epa x assessment	.668		
small programs only	trainee x epa x assessment	.755		

## 4. Discussion

### 4.1. Comparison of large and small residency programs

Overall, we found a broad participation generating a sufficient number of assessments in both large and small residency programs. Considering the voluntary participation, there was a substantial interest in the topic as reflected by a participation rate of 69%, which was even higher in smaller programs than in large ones. This indicates a high motivation to be part in new educational developments and confirms previous results [17], [23], [62]. The very few incomplete data sets might indicate the app's ease of use in daily routine.

In contrast to previous results both in medical education and in the non-healthcare sector, we found no relevant difference between the two sizes of programs. This applies to the number of assessments acquired, the feedback initiated and the distribution over trainees and EPAs, indicating that frequent workplace-based assessments using EPAs are feasible in both settings.

The significantly higher LoS ratings in large programs (indicating less need for supervision) can be due to a variety of reasons: differences in daily routine, expertise, or educational opportunities. For example, in large institutions with longer lasting and higher-risk surgery, arterial cannulation is likely more routine. This could explain why the LoS for this EPA was 0.7 points higher.

On the other hand, there might be a stronger “educational alliance” [63], [64] between trainees and supervisors in small programs, with a higher frequency of one-on-one teaching [41], [43] and less fluctuation of staff [45]. Our finding of a significantly higher proportional number of fixed pairs and assessments acquired by them in small compared to large programs further supports this assumption. Finally, as in non-healthcare industries [49], this could result in a different educational culture in small programs, in which the perception of required and provided supervision might be higher. This could shift the scale of LoS, leading to closer supervision in comparable instances of an EPA. In the end, each individual rating of the LoS by a supervisor is subjective and influenced by the interpersonal relationship [65], [66], [67].

### 4.2. Agreement in LoS

In a previous study using the prEPARED app [23], in 35.6% of assessments there was a divergence in assessed LoS between trainee and supervisor. Similarly, we found this to be the case in 33% of cases and also showed the same clear trend for the trainees to rate themselves slightly less autonomous than their supervisors did.

### 4.3. Required number of assessments and supervisors

We calculated that at least 12 assessments of 3 different EPAs by 4 different supervisors are necessary for a reliable rating in theory (as these ratings had no practical relevance in our setting). The g coefficients  $>0.7$  are in compliance with recommendations for assessments like OSCEs or other high-quality assessments [26], [68]. We even observed higher phi values for small programs, indicating that small programs might at least be comparable to larger ones. Therefore, the EPA assessments via the app used in this study yielded reliable results for random sets of 3 different EPAs out of 10.

When used on a broad, regular basis, it is likely that most residents will quickly achieve sets of more than 3 out of these 10 EPAs, as all of them are expected to be mastered by the end of the second year of residency [56]. Thus, a typical portfolio will display a variety of assessments and support the supervisors' decisions, which clinical tasks can be entrusted to residents. Frequent observations and feedback during daily routine can be expected to improve training as proposed in a programmatic assessment approach. We advise such frequent assessments for improving residence training regardless of the size of the program.

### 4.4. Limitations

A potential source of bias is the voluntary participation of both supervisors and trainees, resulting in a non-representative sample for both groups. Further, it remains open what the practical implications are based on a reliable rating of random sets of 3 different EPAs out of 10. Nonetheless, this data might provide insights for future practice implementation. Investigating the applicability of the findings to mandatory conditions is pivotal. Data on potential confounders such as gender, age, years of training, etc. was not available due to the strict data

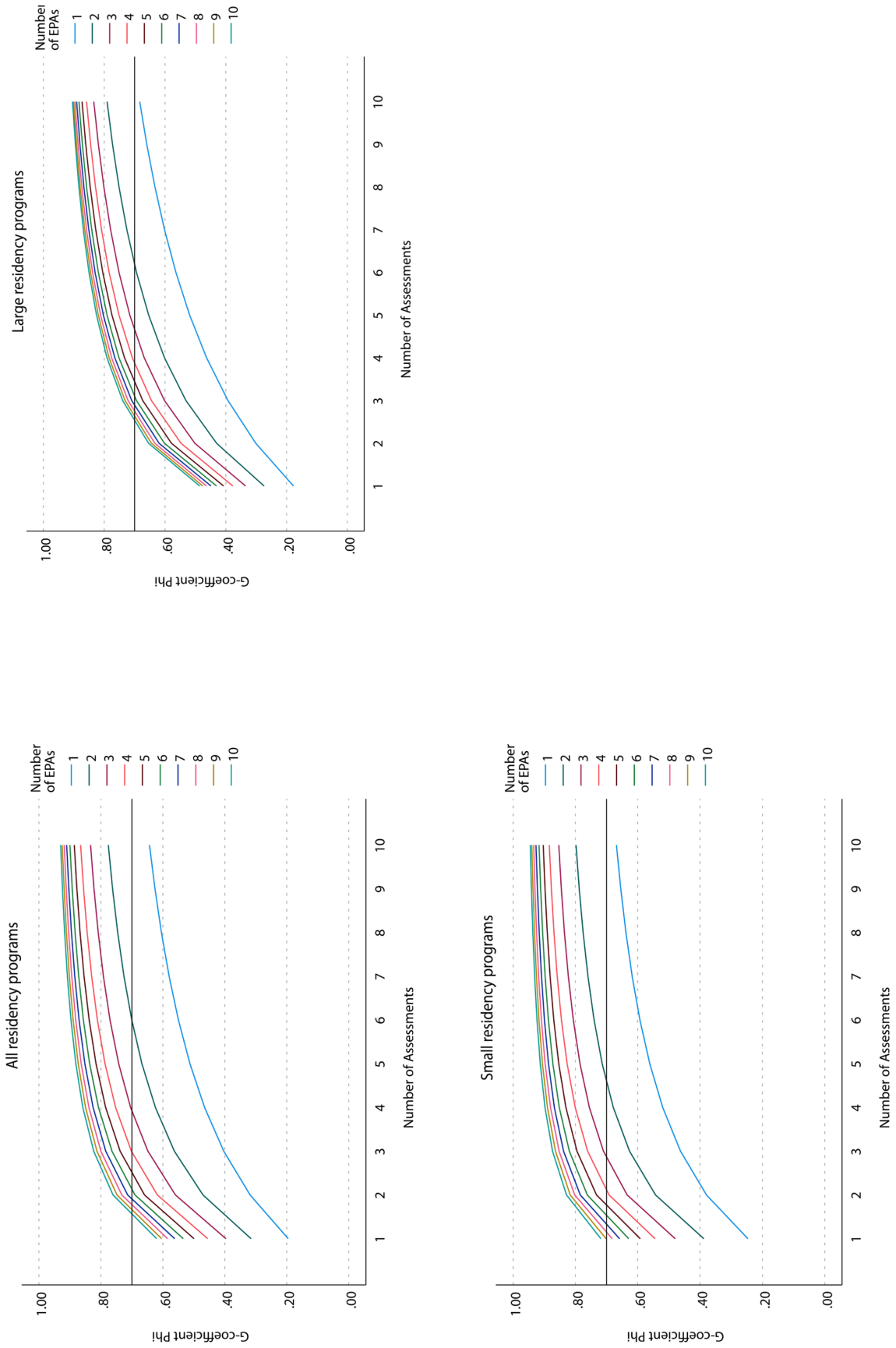


Figure 1: D studies to table 4

privacy policies as stated before. As these factors are likely to have an impact on educational processes, results might have been different if corrected for the confounders. However, this study compared programs of different sizes, and the distribution of confounders is likely to be similar, even more, as rotations in small and large programs are mandatory in Swiss anaesthesiology training.

## 5. Conclusion

Frequent WBAs of EPAs using a mobile application showed to be reliable in both large and small residency programs. We found no relevant differences between the two sizes of programs regarding the numbers and distribution of assessments performed. G theory and D studies analyses confirmed that the minimum number of assessments for a g-coefficient  $>0.7$  can be reached in both program sizes. Therefore, the support of competency-based education through mobile applications to assess EPAs appears suitable in large and small residency programs in anaesthesia.

## Authors' ORCIDs

- Tobias Tessmann: [0000-0002-4687-7452]
- Adrian P. Marty: [0000-0003-3452-9730]
- Daniel Stricker: [0000-0002-7722-0293]
- Sören Huwendiek: [0000-0001-6116-9633]
- Jan Breckwoldt: [0000-0003-1716-1970]

## Competing interests

APM is member of the educational committee of SGAR-SSAR and member of the EPA Committee of the Swiss Institute for Medical Education (SIME). With grant money from the University of Zurich's “Competitive Teaching Grant” and a grant from the SIME, a first functional prototype was developed by an external software company in 2019. In fall 2020, APM founded a company (precisionED Ltd) to rebuild the App from scratch and to provide a sustainable high-quality assessment system. precisionED holds all intellectual property rights and guarantees state-of-the art protection of any data by complying with GDPR-standards.

SH is member of the EPA Committee of the SIME. JB is member of the EPA Committee and co-chair of the “teach the teacher” program of the SIME.

## References

1. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver I, Taber S, Talbot M, Harris KA. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645. DOI: 10.3109/0142159X.2010.501190
2. Park YS, Hodges BD, Tekian A. Evaluating the Paradigm Shift from Time-Based Toward Competency-Based Medical Education: Implications for Curriculum and Assessment. In: Wimmers PF, Mentkowski M, editors. *Assessing Competence in Professional Performance across Disciplines and Professions*. Cham: Springer International Publishing; 2016. p.411-425. DOI: 10.1007/978-3-319-30064-1\_19
3. Jonker G, Hoff RG, Ten Cate OTJ. A case for competency-based anaesthesiology training with entrustable professional activities: an agenda for development and research. *Eur J Anaesthesiol*. 2015;32(2):71-76. DOI: 10.1097/EJA.000000000000109
4. Breckwoldt J, Beckers SK, Breuer G, Marty A. „Entrustable professional activities“: Zukunftsweisendes Konzept für die ärztliche Weiterbildung [Entrustable professional activities: Promising concept in postgraduate medical education]. *Anaesthesist*. 2018;67(6):452-457. DOI: 10.1007/s00101-018-0420-y
5. Englander R, Carraccio C. From theory to practice: making entrustable professional activities come to life in the context of milestones. *Acad Med*. 2014;89(10):1321-1323. DOI: 10.1097/ACM.0000000000000324
6. Ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. 2005;39(12):1176-1177. DOI: 10.1111/j.1365-2929.2005.02341.x
7. Weller JM, Misur M, Nicolson S, Morris J, Ure S, Crossley J, Jolly B. Can I leave the theatre? A key to more reliable workplace-based assessment. *Br J Anaesth*. 2014;112(6):1083-1091. DOI: 10.1093/bja/aeu052
8. Pinilla S, Lerch S, Lüdi R, Neubauer F, Feller S, Stricker D, Berendonk C, Huwendiek S. Entrustment versus performance scale in high-stakes OSCEs: Rater insights and psychometric properties. *Med Teach*. 2023;45(8):885-892. DOI: 10.1080/0142159X.2023.2187683
9. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, Hays R, Palacios Mackay MF, Roberts T, Swanson D. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-1109. DOI: 10.1080/0142159X.2018.1500016
10. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
11. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
12. Van Der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. Twelve Tips for programmatic assessment. *Med Teach*. 2015;37(7):641-646. DOI: 10.3109/0142159X.2014.973388
13. Rich JV, Fostaty Young S, Donnelly C, Hall AK, Dagnone JD, Weersink K, Caudle J, Van Melle E, Klinger DA. Competency-based education calls for programmatic assessment: But what does this look like in practice? *J Eval Clin Pract*. 2020;26(4):1087-1095. DOI: 10.1111/jep.13328
14. Heeneman S, de Jong LH, Dawson LJ, Wilkinson TJ, Ryan A, Tait GR, Rice N, Torre D, Freeman A, van der Vleuten CP. Ottawa 2020 consensus statement for programmatic assessment - 1. Agreement on the principles. *Med Teach*. 2021;43(10):1139-1148. DOI: 10.1080/0142159X.2021.1957088
15. Marty AP, Linsenmeyer M, George B, Young JQ, Breckwoldt J, ten Cate O. Mobile technologies to support workplace-based assessment for entrustment decisions: Guidelines for programs and educators: AMEE Guide No. 154. *Med Teach*. 2023;45(11):1203-1213. DOI: 10.1080/0142159X.2023.2168527



16. Young JQ, Sugarman R, Schwartz J, McClure M, O'Sullivan PS. A mobile app to capture EPA assessment data: Utilizing the consolidated framework for implementation research to identify enablers and barriers to engagement. *Perspect Med Educ.* 2020;9(4):210-219. DOI: 10.1007/s40037-020-00587-z
17. Young JQ, McClure M. Fast, Easy, and Good: Assessing Entrustable Professional Activities in Psychiatry Residents With a Mobile App. *Acad Med.* 2020;95(10):1546-1549. DOI: 10.1097/ACM.0000000000003390
18. George BC, Bohnen JD, Schuller MC, Fryer JP. Using smartphones for trainee performance assessment: A SIMPL case study. *Surgery.* 2020;167(6):903-906. DOI: 10.1016/j.surg.2019.09.011
19. Lefroy J, Roberts N, Molyneux A, Bartlett M, Gay S, McKinley R. Utility of an app-based system to improve feedback following workplace-based assessment. *Int J Med Educ.* 2017;8:207-216. DOI: 10.5116/ijme.5910.dc69
20. Findyartini A, Raharjanti NW, Greviana N, Prajogi GB, Setyorini D. Development of an app-based e-portfolio in postgraduate medical education using Entrustable Professional Activities (EPA) framework: Challenges in a resource-limited setting. *TAPS.* 2021;6(4):92-106. DOI: 10.29060/TAPS.2021-6-4/OA2459
21. Duggan N, Curran VR, Fairbridge NA, Deacon D, Coombs H, Stringer K, Pennell S. Using mobile technology in assessment of entrustable professional activities in undergraduate medical education. *Perspect Med Educ.* 2021;10(6):373-377. DOI: 10.1007/s40037-020-00618-9
22. Diwersi N, Gass JM, Fischer H, Metzger J, Knobe M, Marty AP. Surgery goes EPA (Entrustable Professional Activity) - how a strikingly easy to use app revolutionizes assessments of clinical skills in surgical training. *BMC Med Educ.* 2022;22(1):559. DOI: 10.1186/s12909-022-03622-1
23. Marty AP, Braun J, Schick C, Zalunardo MP, Spahn DR, Breckwoldt J. A mobile application to facilitate implementation of programmatic assessment in anaesthesia training. *Br J Anaesth.* 2022;128(6):990-996. DOI: 10.1016/j.bja.2022.02.038
24. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ.* 2011;45(6):560-569. DOI: 10.1111/j.1365-2923.2010.03913.x
25. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ.* 2012;46(1):28-37. DOI: 10.1111/j.1365-2923.2011.04166.x
26. Peeters MJ, Cor MK. Guidance for high-stakes testing within pharmacy educational assessment. *Curr Pharm Teach Learn.* 2020;12(1):1-4. DOI: 10.1016/j.cptl.2019.10.001
27. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach.* 2012;34(11):960-992. DOI: 10.3109/0142159X.2012.703791
28. Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach.* 2013;35(6):e1197-e1210. DOI: 10.3109/0142159X.2013.788789
29. Andersen SA, Nayahangan LJ, Park YS, Konge L. Use of Generalizability Theory for Exploring Reliability of and Sources of Variance in Assessment of Technical Skills: A Systematic Review and Meta-Analysis. *Acad Med.* 2021;96(11):1609-1619. DOI: 10.1097/ACM.0000000000004150
30. Kreiter C, Zaidi N. Generalizability Theory's Role in Validity Research: Innovative Applications in Health Science Education. *Health Prof Educ.* 2020;6(2):282-290. DOI: 10.1016/j.hpe.2020.02.002
31. Dunne D, Gielissen K, Slade M, Park YS, Green M. WBAs in UME-How Many Are Needed? A Reliability Analysis of 5 AAMC Core EPAs Implemented in the Internal Medicine Clerkship. *J Gen Intern Med.* 2022;37(11):2684-2690. DOI: 10.1007/s11606-021-07151-3
32. Kelleher M, Kinnear B, Sall D, Schumacher D, Schauer DP, Warm EJ, Kelcey B. A Reliability Analysis of Entrustment-Derived Workplace-Based Assessments. *Acad Med.* 2020;95(4):616-622. DOI: 10.1097/ACM.0000000000002997
33. Alves de Lima A, Conde D, Costabel J, Corso J, Van der Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. *Adv Health Sci Educ Theory Pract.* 2013;18(1):5-13. DOI: 10.1007/s10459-011-9343-y
34. Sullivan GM. A Primer on the Validity of Assessment Instruments. *J Grad Med Educ.* 2011;3(2):119-120. DOI: 10.4300/JGME-D-11-00075.1
35. Rourke J. WHO Recommendations to improve retention of rural and remote health workers - important for all countries. *Rural Remote Health.* 2010;10(4):1654. DOI: 10.22605/RRH1654
36. Somporn P, Ash J, Walters L. Stakeholder views of rural community-based medical education: a narrative review of the international literature. *Med Educ.* 2018;52(8):791-802. DOI: 10.1111/medu.13580
37. Health Education England. Training in Smaller Places. London: Health Education England; 2016.
38. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach.* 2001;23(5):445-454. DOI: 10.1080/01421590120075661
39. Boonluksiri P, Thongmak T, Warachit B. Comparison of educational environments in different sized rural hospitals during a longitudinal integrated clerkship in Thailand. *Rural Remote Health.* 2021;21(4):6883. DOI: 10.22605/RRH6883
40. Carmody DF, Jacques A, Denz-Penhey H, Puddey I, Newnham JP. Perceptions by medical students of their educational environment for obstetrics and gynaecology in metropolitan and rural teaching sites. *Med Teach.* 2009;31(12):e596-602. DOI: 10.3109/01421590903193596
41. Parry J, Mathers J, Al-Fares A, Mohammad M, Nandakumar M, Tsivos D. Hostile teaching hospitals and friendly district general hospitals: final year students' views on clinical attachment locations. *Med Educ.* 2002;36(12):1131-1141. DOI: 10.1046/j.1365-2923.2002.01374.x
42. Worley P, Esterman A, Prideaux D. Cohort study of examination performance of undergraduate medical students learning in community settings. *BMJ.* 2004;328(7433):207-209. DOI: 10.1136/bmj.328.7433.207
43. Lyon PM, McLean R, Hyde S, Hendry G. Students' perceptions of clinical attachments across rural and metropolitan settings. *Ass Eval High Educ.* 2008;33(1):63-73. DOI: 10.1080/02602930601122852
44. Kumar N, Brooke A. Should we teach and train in smaller hospitals? *Future Healthc J.* 2020;7(1):8-11. DOI: 10.7861/fhj.2019-0056
45. Bernabeo EC, Holtman MC, Ginsburg S, Rosenbaum JR, Holmboe ES. Lost in Transition: The Experience and Impact of Frequent Changes in the Inpatient Learning Environment. *Acad Med.* 2011;86(5):591-598. DOI: 10.1097/ACM.0b013e318212c2c9
46. Bonnie LHA, Cremers GR, Nasori M, Kramer AWM, van Dijk N. Longitudinal training models for entrusting students with independent patient care?: A systematic review. *Med Educ.* 2022;56(2):159-169. DOI: 10.1111/medu.14607

47. Simpson M, Tuck N, Bellamy S. Small business success factors: the role of education and training. *Educ Train*. 2004;46(8/9):481-491. DOI: 10.1108/00400910410569605
48. Anselmann S. Learning barriers at the workplace: Development and validation of a measurement instrument. *Front Educ*. 2022;7:880778. DOI: 10.3389/educ.2022.880778
49. Ashton D, Sung J, Raddon A, Riordan T. Challenging the myths about learning and training in small and medium-sized enterprises: Implications for public policy. *Employment Working Paper No. 1*. Geneva: International Labour Organization; 2008. p.1:65.
50. Panagiotakopoulos A. Barriers to employee training and learning in small and medium-sized enterprises (SMEs). *Develop Learn Organ*. 2011;25(3):15-18. DOI: 10.1108/14777281111125354
51. Jones P, Beynon MJ, Pickernell D, Packham G. Evaluating the Impact of Different Training Methods on SME Business Performance. *Environ Plann C Gov Policy*. 2013;31(1):56-81. DOI: 10.1068/c12113b
52. Attwell G, Deitmer L. Developing work based personal learning environments in small and medium enterprises. *PLE Conference Proceedings*. 2012;1(1). Zugänglich unter/available from: <https://proa.ua.pt/index.php/ple/article/view/16473>
53. Kock H, Ellström P. Formal and integrated strategies for competence development in SMEs. *J Eur Industr Train*. 2011;35(1):71-88. DOI: 10.1108/030905911111095745
54. Kravčík M, Neulinger K, Klamma R. Boosting Vocational Education and Training in Small Enterprises. In: Verbert K, Sharples M, Klobučar T, editors. *Adaptive and Adaptable Learning*. Cham: Springer International Publishing; 2016. p.600-604. DOI: 10.1007/978-3-319-45153-4\_72
55. Dawe S, Nguyen N. Education and Training that Meets the Needs of Small Business: A Systematic Review of Research. National Centre for Vocational Education Research Ltd; 2007. Zugänglich unter/available from: <https://eric.ed.gov/?id=ED499699>
56. Marty AP, Schmelzer S, Thomasin RA, Braun J, Zalunardo MP, Spahn DR, Breckwoldt J. Agreement between trainees and supervisors on first-year entrustable professional activities for anaesthesia training. *Br J Anaesth*. 2020;125(1):98-103. DOI: 10.1016/j.bja.2020.04.009
57. ten Cate O. Nuts and Bolts of Entrustable Professional Activities. *J Grad Med Educ*. 2013;5(1):157-158. DOI: 10.4300/JGME-D-12-00380.1
58. ten Cate O, Schwartz A, Chen HC. Assessing Trainees and Making Entrustment Decisions: On the Nature and Use of Entrustment-Supervision Scales. *Acad Med*. 2020;95(11):1662-1669. DOI: 10.1097/ACM.0000000000003427
59. ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37(11):983-1002. DOI: 10.3109/0142159X.2015.1060308
60. Brennan R, Norman G. G\_String, A Windows wrapper for urGENOVA©. Hamilton: McMaster Education Research, Innovation & Theory (MERIT); 2017. Zugänglich unter/available from: [http://fhspend.mcmaster.ca/g\\_string/index.html](http://fhspend.mcmaster.ca/g_string/index.html)
61. Hartmann K. Assistentenstellen 2021 pro Fachgebiet, alle Weiterbildungsstätten. 6. Oktober 2021
62. Woodworth GE, Marty AP, Tanaka PP, Ambardekar AP, Chen F, Duncan MJ, Fromer IR, Hallman MR, Klesius LL, Ladlie BL, Mitchel SA, Miller Juve AK, McGrath BJ, Shepler JA, Sims 3rd C, Spofford CM, Van Cleve W, Maniker RB. Development and Pilot Testing of Entrustable Professional Activities for US Anesthesiology Residency Training. *Anesth Analg*. 2021;132(6):1579-1591. DOI: 10.1213/ANE.0000000000005434
63. Telio S, Ajjawi R, Regehr G. The “Educational Alliance” as a Framework for Reconceptualizing Feedback in Medical Education. *Acad Med*. 2015;90(5):609-614. DOI: 10.1097/ACM.0000000000000560
64. Brenner M, Weiss-Breckwoldt AN, Condrau F, Breckwoldt J. Does the ‘Educational Alliance’ conceptualize the student - supervisor relationship when conducting a master thesis in medicine? An interview study. *BMC Med Educ*. 2023;23(1):611. DOI: 10.1186/s12909-023-04593-7
65. Hauer KE, Ten Cate O, Boscardin C, Irby DM, Iobst W, O’Sullivan PS. Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ Theory Pract*. 2014;19(3):435-456. DOI: 10.1007/s10459-013-9474-4
66. Castanelli DJ, Weller JM, Molloy E, Bearman M. Trust, power and learning in workplace-based assessment: The trainee perspective. *Med Educ*. 2022;56(3):280-291. DOI: 10.1111/medu.14631
67. Gille F. About the Essence of Trust: Tell the Truth and Let Me Choose—I Might Trust You. *Int J Public Health*. 2022;67:1604592. DOI: 10.3389/ijph.2022.1604592
68. American Educational Research Association, editor. *Standards for Educational & Psychological Testing*. Washington, D.C: American Educational Research Association; 2014. p.60.

**Corresponding author:**

Prof. Dr. Jan Breckwoldt

University Hospital Zurich, Institute of Anaesthesiology, Raemistr. 100, CH-8091 Zurich, Switzerland  
jan.breckwoldt@usz.ch**Please cite as**

Tessmann T, Marty AP, Stricker D, Huwendiek S, Breckwoldt J. There is no “too small” for frequent workplace-based assessment: Differences between large and small residency programs in anesthesia when using a mobile application to assess EPAs. *GMS J Med Educ*. 2024;41(5):Doc54. DOI: 10.3205/zma001709, URN: urn:nbn:de:0183-zma0017091

**This article is freely available from**<https://doi.org/10.3205/zma001709>**Received:** 2023-10-19**Revised:** 2024-05-08**Accepted:** 2024-06-11**Published:** 2024-11-15**Copyright**

©2024 Tessmann et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

# Es gibt kein „zu klein“ für häufige arbeitsplatzbasierte Assessments: Unterschiede zwischen großen und kleinen Weiterbildungsstätten in der Anästhesie beim Einsatz einer mobilen App zur Bewertung von EPAs

## Zusammenfassung

**Hintergrund:** Kompetenzbasierte Weiterbildung erfordert häufige arbeitsplatzbasierte Assessments (engl. Workplace-based Assessments, WBAs) von anvertraubaren professionellen Tätigkeiten (engl. Entrustable Professional Activities, EPAs). Mobile Apps erhöhen die Effizienz, allerdings ist es nicht bekannt, wie viele Assessments für zuverlässige Bewertungen erforderlich sind und ob das Konzept in Facharzt-Weiterbildungsstätten jeder Größe umgesetzt werden kann.

**Methoden:** Über 5 Monate wurden mit einer Smartphone-App 10 verschiedene EPAs im klinischen Alltag in Schweizer Anästhesieabteilungen beurteilt. Die Daten von großen und kleineren Weiterbildungsstätten wurden verglichen. Mittels Generalisierbarkeitstheorie wurde eine Mindestanzahl an Assessments berechnet, die für zuverlässige Beurteilungen erforderlich sind.

**Ergebnisse:** In 28 Weiterbildungsstätten wurden 3936 Assessments von 306 Supervidierenden für 295 Trainees erfasst. Die mediane Anzahl an Assessments pro Trainee lag bei 8, mit einem Median von 4 verschiedenen EPAs, die von 3 verschiedenen Supervidierenden beurteilt wurden. Wir fanden keine statistisch signifikanten Unterschiede zwischen großen und kleinen Weiterbildungsstätten in Bezug auf die Anzahl der Assessments pro Trainee, pro Supervisor\*in, pro EPA, die Übereinstimmung zwischen Supervidierenden und Trainees oder die Anzahl der angeregten Feedbackprozesse. Der durchschnittliche Grad an Supervision (engl. *level of supervision*, LoS, Skala von 1 bis 5) lag bei größeren Programmen bei 3,2 (SD 0,5) im Vergleich zu 2,7 (SD 0,4) ( $p < 0,05$ ). Um einen  $g$ -Koeffizienten  $> 0,7$  zu erreichen, musste mindestens eine zufällige Auswahl von 3 verschiedenen EPAs bewertet werden, wobei jede EPA mindestens viermal von vier verschiedenen Supervidierenden bewertet werden musste, was insgesamt 12 Assessments ergab.

**Schlussfolgerung:** Häufige WBAs von EPAs waren sowohl in großen als auch in kleinen Weiterbildungsstätten durchführbar. Wir fanden keine signifikanten Unterschiede in der Anzahl der durchgeführten Assessments. Die für einen  $g$ -Koeffizienten  $> 0,7$  erforderliche Mindestanzahl von Assessments war in großen und kleinen Weiterbildungsstätten erreichbar.

**Schlüsselwörter:** kompetenzbasierte Weiterbildung, CBME, arbeitsplatzbasierte Assessments, WBA, anvertraubare professionelle Tätigkeiten, EPA, entrustment, Feedback, Entscheidungsfindung, Facharztweiterbildung, Generalisierbarkeitstheorie,  $g$ -Theorie, Reliabilität, kleine Unternehmen, KMU, Smartphone, mobile App

Tobias Tessmann<sup>1</sup>

Adrian P. Marty<sup>2</sup>

Daniel Stricker<sup>3</sup>

Sören Huwendiek<sup>3</sup>

Jan Breckwoldt<sup>1</sup>

1 Universitätsspital Zürich, Institut für Anästhesiologie, Zürich, Schweiz

2 Universitätsklinik Balgrist, Abteilung für Anästhesiologie, Intensivmedizin und Schmerztherapie, Zürich, Schweiz

3 Universität Bern, Institut für medizinische Ausbildung, Bern, Schweiz

# 1. Einleitung

## 1.1. Kompetenzbasierte medizinische Ausbildung (engl. Competency-based medical education, CBME)

CBME und ihre Umsetzung durch das Konzept der „Anvertraubaren professionellen Tätigkeiten“ (engl. Entrustable Professional Activities, EPAs) findet in der medizinischen Weiterbildung zunehmend Anklang [1], [2]. Eine EPA kann als ein Rahmen gesehen werden, durch den eine oder mehrere Kompetenzen während umschriebener, spezifischer Aufgaben in der klinischen Praxis beobachtet werden können [3], [4]. Die Summe aller EPAs eines Fachgebiets repräsentiert dessen gesamtes Curriculum und damit alle geforderten Kompetenzen zu einem bestimmten Zeitpunkt der Weiterbildung. Wenn sie in einem Portfolio kombiniert werden, können bestimmte Gruppen komplementärer EPAs wichtige Entwicklungsstufen [5] in einem ganzheitlichen, longitudinalen Ansatz abbilden. Das erstmals von Ten Cate 2005 [6] veröffentlichte Konzept der EPAs ergänzte und erleichterte nicht nur den Aufstieg der CBME, sondern führte auch den wichtigen Aspekt des Vertrauens und der expliziten Entscheidung ein, jemandem zukünftig etwas anzuvertrauen (engl. *entrustment*). Es ist erwiesen, dass die Verwendung von *Entrustment-Supervision*-Skalen anstelle von numerischen Skalen eine größere klinische Relevanz, eine bessere Inter-Rater-Reliabilität und eine einfachere Anwendbarkeit aufweist [7], [8]. Sie liefern keine einfache Schulnote, sondern eine Antwort auf die Frage: Wie viel Aufsicht wird der/die Trainee in einer ähnlichen Situation beim nächsten Mal benötigen?

Zur Beurteilung der EPA wird eine Palette verschiedener arbeitsplatzbezogener Assessments (engl. *workplace-based assessment*, WBA) eingesetzt. Jedes WBA sollte nicht nur als Instrument zur Bewertung des aktuellen Status der Trainees dienen, sondern auch ihre zukünftige Entwicklung durch häufiges Feedback unterstützen [9]. Dieses Konzept des „Assessments für das Lernen“ [10] wurde als eine der Schlüsselkomponenten eines *programmatic assessments* eingeführt. Darüber hinaus sollten die Ergebnisse dieser häufigen WBAs nicht einzeln für gravierende Entscheidungen (engl. *high-stake decisions*) wie das „Bestehen/Nichtbestehen“ eines Weiterbildungsjahres verwendet werden. Vielmehr sollten diese Entscheidungen in so genannten Kompetenzausschüssen auf der Grundlage von vielen Daten aus verschiedenen Quellen [11], [12], [13], [14] getroffen werden.

Um häufigere und kürzere Assessments durchführen zu können, wurden mobile Apps entwickelt [15], [16], [17], [18], [19], [20], [21], [22], die den Zeitaufwand für ein einzelnes Assessment erheblich reduzieren [23]. Dennoch ist eine beträchtliche Anzahl an Assessments erforderlich, um das Kompetenzniveau eines/einer Trainee zuverlässig einschätzen zu können. Reliabilität und Validität eines Instruments hängen eng mit seinem Design und seiner Verwendung im entsprechenden Kontext zusammen [24],

[25], [26], [27], [28], [29]. Unabhängig davon, zu welchem Zweck ein Assessment durchgeführt wird und zu welchen Entscheidungen seine Ergebnisse beitragen, müssen diese auf zuverlässigen, transparenten und korrekten Informationen basieren.

Die Verwendung von *Entrustment-Supervision*-Skalen ist nicht nur mit einem geringeren Arbeitsaufwand für die Bewertenden verbunden, sondern reduziert auch die erforderliche Anzahl an Assessments um etwa 50% für Mini-CEX [7]. Die publizierten Mindestzahlen anderer WBA-Instrumente liegen jedoch in einem weiten Bereich [24], [29], [30], [31], [32], [33], und die Mindestanzahl an Assessments für ein App-basiertes WBA von EPAs ist völlig offen. Die Mindestanzahl der Assessments könnte von einer Vielzahl von Faktoren abhängen, z. B. von der Anzahl der verschiedenen Prüfenden, der Interaktion zwischen Supervidierenden und Trainees oder der Größe einer Weiterbildungsstätte [4], [8], [25], [34].

## 1.2. Unterschiede zwischen großen und kleinen Weiterbildungsstätten

Kleinere Einrichtungen des Gesundheitswesens spielen eine entscheidende Rolle für einen Zugang aller Menschen zur Gesundheitsversorgung weltweit, insbesondere in ländlichen und städtischen unterversorgten Gebieten [35], [36]. Entsprechend ist eine qualitativ hochwertige medizinische Weiterbildung in kleineren Einrichtungen unabdingbar. Neben ihren strukturellen Bedingungen (Anzahl der Trainees und Supervidierenden, Fallschwere und -volumen, verfügbare Ressourcen) [37], unterscheiden sich kleinere Einrichtungen in ihrer Ausbildungskultur [38]. Es ist aber keine Frage, eines „besser oder schlechter“, sondern der unterschiedlichen Möglichkeiten: Während eine Untersuchung feststellte, dass große Weiterbildungsstätten von der besseren Infrastruktur ihrer Einrichtungen profitieren [39], berichteten andere Autoren über keine signifikanten Unterschiede [40], oder stellten gar fest, dass kleinere Einrichtungen bessere Möglichkeiten zur Entwicklung klinischer Fertigkeiten, eine bessere Haltung der Supervidierenden und ein höheres Gefühl der Selbstwirksamkeit in der medizinischen Basisausbildung bieten [41], [42], [43]. Die stärker spezialisierte Versorgung in großen Einrichtungen wird durch einen generalistischeren und ganzheitlicheren Ansatz in kleineren Einrichtungen aufgewogen. Darüber hinaus bieten kleinere Weiterbildungsstätten möglicherweise bessere Chancen für die Entwicklung nicht-klinischer Fähigkeiten wie Belastbarkeit und Autonomie, sind aber gleichzeitig diesbezüglich anspruchsvoller [44]. Die vertrautere Atmosphäre, länger anhaltende Mentoring-Beziehungen und eine geringere Personalfuktuation [45] könnten sich positiv auf die Qualität der Ausbildung auswirken [46].



### 1.3. Analogien außerhalb des Gesundheitswesens

Außerhalb des Gesundheitswesens hat sich gezeigt, dass kleine und mittlere Unternehmen (KMU) ihre eigenen Bedürfnisse, Schwierigkeiten und Merkmale in Bezug auf die arbeitsplatzbezogene Ausbildung haben, was zu einer spezifischen Ausbildungskultur führt [47], [48], [49], [50]. Sie kann in zweierlei Hinsicht charakterisiert werden: Sie ist eher eine integrierte als eine formalisierte Ausbildung und findet in einer Lernumgebung statt, die eher fördert als einschränkt [51], [52]. Wenn die unterschiedliche Ausbildungskultur in KMU aufgenommen wird, führt das zu besseren Ergebnissen [53], [54]. In einem Bericht heißt es: „Strategien, die der Lernweise eines kleinen Unternehmens entsprechen, sind erfolgreicher als formale Schulungen“ [55].

### 1.4. Forschungsfrage

Häufige WBAs von EPAs in Weiterbildungsstätten unterschiedlicher Größe wurden bisher nicht untersucht. Erkenntnisse über die Unterschiede und Gemeinsamkeiten zwischen kleineren und größeren Weiterbildungsstätten könnten helfen zu klären, ob häufig durchgeführte Assessments zuverlässig über das ganze Spektrum der Weiterbildungsstätten angewandt werden können.

Wir haben daher die Nutzung und die Ergebnisse des Assessments von EPAs mit einer mobilen App zwischen großen und kleinen Weiterbildungsstätten für Anästhesiologie verglichen. Zusätzlich haben wir eine Reliabilitätsanalyse durchgeführt, um die minimal erforderliche Anzahl von Assessments und Supervidierenden für zuverlässige Bewertungen zu ermitteln.

## 2. Methoden

### 2.1. Teilnehmer und Kategorisierung der Weiterbildungsstätten

Die Ausbildungskommission der Schweizerischen Gesellschaft für Anästhesiologie (SGAR-SSAR) lud alle Schweizer Anästhesieabteilungen mit Weiterbildungsstätten ein, sich (kostenlos) zu beteiligen. Während 5 Monaten des Jahres 2021 wurden alle mittels der mobilen App aufgezeichneten Assessments analysiert. Abteilungen mit Weiterbildungsberechtigung für 3,5 oder 3 Jahren (nationale Kategorien A1 und A2) definierten wir als „große Weiterbildungsstätte“ und solche für 1 oder 2 Jahren (nationale Kategorien C und B) als „kleine Weiterbildungsstätte“. Die großen Abteilungen verfügen über einen Lehrkörper von mehr als 50 Personen und gehören zu Spitälern, die in einer Vielzahl von Fachgebieten eine Maximalversorgung anbieten. Die kleineren Abteilungen repräsentieren das weite Spektrum der Grundversorgung und weisen die meisten der für die obige Klassifizierung als „klein“ verwendeten Merkmale auf [37]. Für den Ab-

schluss der Facharztweiterbildung ist in der Schweiz eine Mindestzeit von einem Jahr in einer „kleinen“ Einrichtung erforderlich.

### 2.2. Mobile Smartphone-App und EPAs

Die Smartphone-App „prEPared“ (*precisionED Ltd, Wollerau, Schweiz*) ermöglicht die Bewertung von EPAs nahezu in Echtzeit mit minimaler Unterbrechung des Arbeitsablaufs [23]. Der/die Trainee initiiert den Prozess, indem die durchgeführte EPA ausgewählt wird. Anschließend bewerten sowohl der/die Trainee als auch der/die Supervisor\*in unabhängig voneinander die Komplexität der EPA („einfach“ vs. „komplex“) sowie den eingeschätzten Grad an erforderlicher Supervision (engl. *level of supervision, LoS*) für einen gleichgelagerten Fall in der Zukunft. Der Vergleich der Bewertungen regt dann ein Feedbackgespräch an, das optional auch durch strukturierte Feedbackanweisungen von der App geleitet werden kann. Jedes Assessment wird im ePortfolio des/der Trainee gespeichert, das auch mit den Supervidierenden geteilt werden kann [23].

Die Einführung der prEPared-App wurde mit Informationsmaterial wie Lehrvideos über CBME, EPAs und die App begleitet.

Für den Studien-Zeitraum verwendeten wir zehn gängige EPAs (siehe Tabelle 1, linke Spalte), die weitgehend den Katalog des ersten bis zweiten Jahres des nationalen Weiterbildungscurriculums der Anästhesie [56] abbilden. Dennoch stand der gesamte EPA-Katalog in der App zur Auswahl (um den Trainees und Supervidierenden den größten Nutzen für die individuelle Ausbildung zu bieten). Die Level of Supervision reichten von 1 (=nur beobachten) bis 5 (=andere supervidieren) [57], [58], [59], wie in Tabelle 2 dargestellt.

### 2.3. Statistik und G-Theorie-Modell

Als primärer Endpunkt wurde die Anzahl der Assessments pro Trainee und pro Supervisor\*in verglichen. Sekundäre Endpunkte waren Unterschiede für die einzelnen bewerteten EPAs, die jeweiligen Bewertungen (erforderliche LoS gemäss Supervisor\*in), die Übereinstimmung im LoS von Trainee und Supervidierenden, die Häufigkeit des durch die Bewertung angeregten Feedbacks und schließlich eine Reliabilitätsanalyse, die die minimal erforderliche Anzahl von Assessments und Supervidierenden angibt, um zuverlässige Bewertungen eines/einer Trainee in einem theoretischen Modell zu erzielen (unter Verwendung der G-Theorie und D-Studien).

Für den allgemeinen Vergleich der Bewertungen von Supervidierenden größerer und kleinerer Weiterbildungsstätten wurden Wilcoxon-U-Tests durchgeführt. Bei Mehrfachtests wurde eine Bonferroni-Korrektur des Alpha-Fehlers vorgenommen.

Die G-Theorie wird routinemäßig für die Evaluation von Bewertungsinstrumenten verwendet, indem versucht wird, die Varianzbeträge für einzelne Facetten und ihre



**Tabelle 1: Einzelne EPAs mit der Anzahl n Assessments und den durchschnittlichen Bewertungen pro Kategorie, in absteigender Reihenfolge der Gesamtanzahl**

EPA-Name	Messung	Kategorie Weiterbildungsstätte			Unterschied (groß-klein)	p (2-seitig U-Test)
		Ins-gesamt	Groß	Klein		
Einleitung einer Allgemeinanästhesie	n Assessments	947	778	169		
	durchschn. Bewertung	2.98	3.05	2.66	0.39	< .001 *
	Übereinstimmung der Bewertung	0.12	0.13	0.04	0.09	.160
Endotracheale Intubation	n Assessments	563	425	138		
	durchschn. Bewertung	2.86	2.93	2.64	0.3	< .001 *
	Übereinstimmung der Bewertung	0.1	0.13	-0.02	0.15	.025
Intraoperatives Management	n Assessments	438	367	71		
	durchschn. Bewertung	3.13	3.28	2.36	0.92	< .001 *
	Übereinstimmung der Bewertung	0.02	0.03	-0.05	0.07	.202
Spinalanästhesie	n Assessments	410	312	98		
	durchschn. Bewertung	3.06	3.14	2.81	0.33	< .001 *
	Übereinstimmung der Bewertung	0.06	0.05	0.08	-0.03	.728
Aufwachen aus der Allgemeinanästhesie	n Assessments	333	269	64		
	durchschn. Bewertung	3.08	3.16	2.71	0.45	< .001 *
	Übereinstimmung der Bewertung	0.1	0.12	0.02	0.1	.192
Präoperative Beurteilung/Planung/ Zustimmung	n Assessments	265	216	49		
	durchschn. Bewertung	3.55	3.66	3.08	0.58	< .001 *
	Übereinstimmung der Bewertung	-0.04	-0.02	-0.13	0.11	.247
Platzierung der Arterienleitung (radial)	n Assessments	253	215	38		
	durchschn. Bewertung	3.51	3.62	2.89	0.73	< .001 *
	Übereinstimmung der Bewertung	0.07	0.06	0.14	-0.08	.468
Axillarer Plexus brachialis Block	n Assessments	250	207	43		
	durchschn. Bewertung	2.85	2.89	2.7	0.19	.457
	Übereinstimmung der Bewertung	0.08	0.04	0.24	-0.2	.033
Verschiedenes. EPAs	n Assessments	166	90	76		
	durchschn. Bewertung	2.57	2.98	2.09	0.89	< .001 *
	Übereinstimmung der Bewertung	0.07	0.09	0.04	0.05	.588
Einsetzen der Larynxmaske	n Assessments	156	100	56		
	durchschn. Bewertung	3.06	3.23	2.75	0.48	.001 *
	Übereinstimmung der Bewertung	0.05	0.07	0.02	0.06	.768
Übergabe	n Assessments	155	107	48		
	durchschn. Bewertung	3.59	3.8	3.1	0.7	< .001 *
	Übereinstimmung der Bewertung	-0.13	-0.1	-0.21	0.12	.388
Summen	n Assessments	3936	3086	850		
	durchschn. Bewertung	3.07	3.18	2.67	0.51	< .001 *
	Übereinstimmung der Bewertung	0.06	0.08	0.01	0.06	.019

Tabelle 2: Beschreibung der Entrustment-Supervision-Skala

LoS	Beschreibung	Beschreibung nach [58]
5	Andere supervidieren	Kann als Supervisor*in jüngerer Trainees für diese EPA fungieren.
4	Entfernte Supervision	Kann EPA unbeaufsichtigt praktizieren (unter Fernaufsicht). Aufsichtspersonen können per Telefon usw. erreichbar sein.
3	Indirekte Supervision	Kann EPA unter indirekter (reaktiver) Supervision ausüben. Die Supervidierenden sind nicht unbedingt im Raum, aber schnell verfügbar.
2	Direkte Supervision	Darf EPA unter direkter (proaktiver) Supervision praktizieren. Die Supervidierenden befinden sich in der Regel im selben Raum.
1	Nur Beobachten	Darf anwesend sein, aber nicht die EPA ausführen.

Wechselwirkungen zu quantifizieren und zuzuordnen [27], [29].

Um einen ausgewogenen Datensatz für die Berechnung der G-Theorie zu erhalten, wählten wir aus allen möglichen Trainee/Supervisor\*in-Kombinationen eine Stichprobe von 20 Trainees aus, die jeweils 3 EPAs mit 4 Assessments durch 4 verschiedene Supervidierende abgeschlossen hatten. Diese Entscheidung wurde a priori für die statistische Analyse getroffen, basierend auf den Literaturergebnissen bezüglich der Mindestanzahl für ähnliche Beurteilungsinstrumente [29], [30], [31], [32], [33]. 16 der 20 Trainees wurden in großen Weiterbildungsstätten ausgebildet, die anderen 4 in kleinen Weiterbildungsstätten, was proportional die Gesamtstichprobe widerspiegelt.

Wir berechneten den Einfluss der einzelnen EPA, der Bewertung (erforderliche LoS gemäss Supervisor\*in) und der Weiterbildungsstätte auf die Varianz der Bewertungen der Trainees. Das Modell der G-Theorie lautete: EPA x Bewertung x (Trainee : Weiterbildungsstätte).

Die jeweiligen Varianzkomponenten wurden in Entscheidungs-Studien (D-Studien) verwendet, um die minimale Anzahl von Assessments bzw. Supervidierenden zu bestimmen.

Eine Generalisierbarkeits-Studie (EPA x Assessment x Trainee) und D-Studien wurden auch separat für größere und kleinere Weiterbildungsstätten berechnet, um sie miteinander und mit der Gesamtberechnung zu vergleichen. Zur Kontrolle und Quantifizierung möglicher Unterschiede und Interaktionseffekte wurde eine dreifaktorielle Varianzanalyse mit teilweiser Messwiederholung mit der Kategorie der Weiterbildungsstätte (groß vs. klein) als Faktor zwischen den Gruppen und EPA (3 Stufen) sowie Supervisor\*in (4 Stufen) als Faktoren innerhalb der Trainees durchgeführt. Die abhängige Variable war der LoS der Supervidierenden. Nur die 20 Trainees, die für die Analyse der g-Studie in Frage kamen, wurden für diese Analyse berücksichtigt.

G-Theorie und D-Studien wurden mit G\_String A Windows Wrapper for urGENOVA [60] berechnet. Alle anderen statistischen Berechnungen wurden mit SPSS für Windows Version 28 (IBM, Armonk, NY, USA) durchgeführt.

## 2.4. Datensicherheit und Ethik

Aufgrund des strengen Datenschutzes der precisionED AG und um Bedenken bei der Teilnahme zu vermeiden,

wurden keine weiteren Daten (z.B. Alter, Geschlecht, Ausbildungsjahre) erhoben. Die Studie wurde von der Ethik-Kommission des Kantons Zürich unter BASEC-Nr. Req-2019-00242 (21. März 2019) freigegeben.

## 3. Ergebnisse

### 3.1. Deskriptive Zusammenfassung der Daten

Zwischen dem 1. April und dem 31. August 2021 nutzten 306 Supervidierende und 295 Trainees die prEPARED-App, um 3936 Assessments zu erfassen. Wir schlossen 93 unvollständige Assessments (durch Trainee und Supervisor\*in, 2,3% der Gesamtstichprobe) im Vorfeld aus dem Datensatz aus. Außerdem sowie 197 teilweise unvollständige Assessments (entweder durch Trainee oder Supervisor\*in, 4,5% der Gesamtstichprobe) von den jeweiligen Analysen ausgeschlossen.

Insgesamt nahmen 28 der 53 Weiterbildungsstätten (52,8%) teil, davon 15 große und 13 kleine. Rund drei Viertel der Supervidierenden und Trainees arbeiteten in großen Programmen und dementsprechend wurden dort auch drei Viertel aller Assessments erfasst.

Von 722 registrierten Weiterbildungsassistent\*innen in Anästhesie in der Schweiz im Jahr 2021 [61], nahmen 41% an der Studie teil. In den an der Studie teilnehmenden Weiterbildungsstätten waren 496 gemeldet, was einer Teilnahmequote von 69% entspricht (58% in großen, 78% in kleinen Weiterbildungsstätten).

Die durchschnittliche Anzahl der Assessments pro Trainee betrug 8, die durchschnittliche Anzahl unterschiedlicher EPAs 4. Im Durchschnitt wurde ein\*e Trainee von 3 verschiedenen Supervidierenden beurteilt. In 2261 (57 %) aller Assessments wurde ein Feedback-Prozess begonnen und in 2165 (96 %) davon wurde ein Lernziel dokumentiert (siehe Tabelle 3).

### 3.2. Vergleich der Ergebnisse von großen und kleinen Weiterbildungsstätten

Insgesamt konnten wir keine relevanten Unterschiede zwischen großen und kleinen Weiterbildungsstätten feststellen. Insbesondere fanden wir keinen signifikanten Unterschied bei der Anzahl der Assessments, der Trai-

Tabelle 3: Anzahl der Assessments, Trainees, Supervidierenden und jeweilige Durchschnittswerte pro Kategorie

Kategorie Weiterbildungs- stätte	Gesamt				Groß				Klein			
	n	%	Median	IQR	n	%	Median	IQR	n	%	Median	IQR
Assessments	3936	100	49	112.5	3086	78.4	85	265.5	850	21.6	45	81
...mit Feedbackprozess	2261	57.4	40	87.5	1817	58.9	77	122.5	444	52.2	28	32
Trainees	295	100	4	9	253	85.8	10	24	45	15.3	3	1
Supervidierende	306	100	7	8.3	223	72.9	9	18	85	27.8	5	4
Durchschnittliche Supervidierende pro Trainee			3				3				3	
Durchschnittliche Assessments pro Trainee			8				8				8	
Durchschnittliche Assessments pro Supervisor*in			7				8				5	

nees, der Supervidierenden, der durchschnittlichen Anzahl der Supervidierenden pro Trainee, der durchschnittlichen Anzahl der Assessments pro Trainee oder pro Supervisor\*in oder der Häufigkeit des angeregten Feedbacks (siehe Tabelle 3). Wir fanden auch keine statistisch signifikanten Unterschiede in der Übereinstimmung zwischen Supervidierenden und Trainees hinsichtlich des notwendigen LoS oder der Komplexität der EPA. Außerdem war die Verteilung der einzelnen EPAs zwischen den Gruppen ähnlich (siehe Tabelle 1), ebenso wie die durchschnittliche Bewertung der Komplexität.

Ein statistischer Unterschied ( $p < 0,01$ ) war nachweisbar für die durchschnittlichen LoS-Bewertungen der Supervidierenden, die in großen Programmen für fast alle EPAs durchschnittlich 0,5 Punkte höher ausfielen. Ein weiterer signifikanter Unterschied zwischen den Gruppen war die höhere Anzahl „fester Paare“ in kleineren Weiterbildungsstätten, vgl. Abschnitt 3.4.

### 3.3. Details zu den EPAs und den jeweiligen Bewertungen

Die „Einleitung der Anästhesie“ und die „endotracheale Intubation“ waren die beiden am häufigsten bewerteten EPAs und machten ein Drittel aller Assessments aus. Die übrigen Assessments verteilten sich gleichmäßig auf die anderen EPAs (vgl. Tabelle 1).

Es gab keine relevanten Unterschiede in den LoS-Bewertungen zwischen den einzelnen EPAs, mit einem Gesamtdurchschnitt von 3,11 (SD 0,32), d.h. es ist eine indirekte Supervision erforderlich. Der von den Trainees selbst eingeschätzte LoS war für alle EPAs etwas niedriger als die Bewertungen der Supervidierenden, aber es konnte kein statistisch signifikanter Unterschied nachgewiesen werden. Das Gleiche gilt für die Bewertung der Komplexität, wobei im Mittel etwa ein Drittel aller bewerteten Ereignisse als komplex eingestuft wurde.

### 3.4. „Feste Paare“ und „Enthusiasten“

In beiden Größenkategorien fanden wir feste Paare (d. h. ein Trainee, der/die mit dem/der gleichen Supervisor\*in für 4 oder mehr Assessments zusammenarbeitete). Von diesen 317 Paaren arbeiteten 250 (22,4% von 1116) bzw. 67 (38,3% von 175) in großen bzw. kleinen Weiterbildungsstätten. Der Unterschied war statistisch signifikant ( $\chi^2(1)=20,604$ ,  $p < 0,001$ ). Auf diese Paare entfielen 1843 (46,8% der Gesamtzahl) aller Assessments: 1381 (44,7%) in großen und 462 (52,7%) in kleinen Weiterbildungsstätten ( $\chi^2(1)=24,68$ ;  $p < 0,001$ ).

Bei den Trainees und den Supervidierenden variierte die Anzahl der Assessments stark, wobei einzelne Personen mehr als 200 Assessments abgegeben haben (2 Supervidierende und 1 Trainee aus großen und 1 bzw. 2 aus kleinen Weiterbildungsstätten). 3 der 15 großen und 5 der 13 kleinen Weiterbildungsstätten gaben nur wenige Assessments ab (<34 Assessments).

### 3.5. G-Theorie

Die Ergebnisse für die G-Theorie-Modelle sind in Tabelle 4 dargestellt. Sie ergaben ein robustes Ergebnis von  $\phi > 0,7$  für die gewählte Kombination von Assessments/EPAs/Supervidierenden, während 21% der Varianz der interessierenden Facette (dem Trainee) zugeschrieben wurden, eingebunden in die Weiterbildungsstätten (wie vom Studiendesign vorgegeben). Für die 20 Trainees, die in der G-Theorie-Analyse eingeschlossen waren, ergab die 3-Wege-ANOVA keine signifikanten Unterschiede in den LoS-Bewertungen zwischen Weiterbildungsstätten, EPAs oder Supervidierenden, und es gab auch keine signifikanten Interaktionen.

Alle entsprechenden D-Studien sind in Abbildung 1 grafisch dargestellt. Das Hauptergebnis ist in Abbildung 1 zu finden (dunkelrote Linie): Theoretisch zufriedenstellende ( $\phi$ -Koeffizient von 0,7) Assessments sind mit einer zufälligen Auswahl von mindestens 3 verschiedenen EPAs

Tabelle 4: Ergebnisse der G-Theorie für 3 EPAs, mit je 4 Assessments von 4 verschiedenen Supervidierenden

Kategorie Weiterbildungsstätte	Generalisierbarkeitsmodell	Phi	Prozentsatz der Varianz Weiterbildungsstätte	Prozentsatz der Varianz Trainee: Weiterbildungsstätte
Alle	(Trainee: Weiterbildungsstätte) x Epa x Assessment	.704	0%	21%
Nur große Weiterbildungsstätten	Trainee x epa x Assessment	.668		
Nur kleine Weiterbildungsstätten	Trainee x epa x Assessment	.755		

aus 10 zu erwarten, wobei jedes EPA mindestens viermal von vier verschiedenen Supervidierenden bewertet wird, was insgesamt 12 Bewertungen ergibt.

Ein Phi-Koeffizient von 0,8, der als außerordentlich robustes Messinstrument interpretiert werden kann, ist zu erwarten, wenn zufällige Sätze von 6 EPAs mit 4 verschiedenen Supervidierenden (wiederum insgesamt 12 Bewertungen) bewertet werden.

## 4. Diskussion

### 4.1. Vergleich zwischen großen und kleinen Weiterbildungsstätten

Insgesamt stellten wir eine breite Beteiligung fest, die zu einer ausreichenden Anzahl an Assessments sowohl in großen als auch in kleinen Weiterbildungsstätten führte. Bedenkt man die freiwillige Teilnahme, bestand ein erhebliches Interesse an dem Thema, was sich in einer Teilnahmequote von 69% widerspiegelt, die in kleineren Programmen noch höher war als in großen Programmen. Dies deutet auf eine hohe Motivation hin, Teil von neuen Entwicklungen in der Weiterbildung zu sein, und bestätigt frühere Ergebnisse [17], [23], [62]. Die wenigen unvollständigen Datensätze könnten auf eine einfache Nutzbarkeit der App in der täglichen Routine hinweisen.

Im Gegensatz zu früheren Ergebnissen sowohl in der medizinischen Weiterbildung als auch außerhalb des Gesundheitswesens fanden wir keinen relevanten Unterschied zwischen den beiden Größen der Weiterbildungsstätten. Dies gilt für die Anzahl der erfassten Assessments, die angestoßenen Feedback-Prozesse und die ähnliche Verteilung über Trainees wie EPAs, und zeigt auf, dass häufige arbeitsplatzbezogene Assessments unter Verwendung von EPAs in beiden Lern-Umfeldern umsetzbar sind.

Die signifikant höheren LoS-Bewertungen in großen Weiterbildungsstätten (die auf einen geringeren Bedarf an Supervision hinweisen) können auf eine Vielzahl von Gründen zurückzuführen sein: Unterschiede in der täglichen Routine, im Fachwissen oder der Ausbildungsmöglichkeiten. In großen Einrichtungen mit länger andauernden und risikoreicheren Eingriffen beispielsweise ist die arterielle Kanülierung wahrscheinlich eher Routine. Dies

könnte erklären, warum der LoS für diese EPA um 0,7 Punkte höher war.

Andererseits könnte es in kleinen Weiterbildungsstätten eine stärkere „*educational alliance*“ [63], [64] zwischen Supervidierenden und Trainees geben, mit einer höheren Häufigkeit von 1-zu-1-Betreuung [41], [43] und einer geringeren Personalfuktuation [45]. Unser Ergebnis, dass die Anzahl der festen Paare und der von ihnen durchgeführten Assessments in kleinen Weiterbildungsstätten signifikant höher ist als in großen, unterstützt diese Annahme zusätzlich. Schließlich könnte dies, wie in anderen Branchen außerhalb des Gesundheitswesens [49], zu einer anderen Ausbildungskultur in kleinen Weiterbildungsstätten führen, in der die Wahrnehmung der erforderlichen und gebotenen Supervision höher sein könnte. Dies könnte die Skala des LoS verschieben und zu einer engeren Supervision in vergleichbaren Fällen einer EPA führen. Letztendlich ist jede individuelle Bewertung des LoS durch Supervidierende subjektiv und von der interpersonellen Beziehung beeinflusst [65], [66], [67].

### 4.2. Übereinstimmung im LoS

In einer früheren Studie, in der die prEPared-App [23] verwendet wurde, gab es in 35,6% der Assessments eine Divergenz der Bewertung des LoS zwischen Trainees und Supervidierenden. Wir bestätigten, dies in unserer Studie mit 33% der Fälle, und es zeigte sich auch die gleiche Tendenz, dass Trainees sich selbst etwas weniger selbstständig einschätzten als ihre Supervidierenden.

### 4.3. Erforderliche Anzahl von Assessments und Supervidierenden

Wir konnten errechnen, dass in der Theorie mindestens 12 Assessments von 3 verschiedene EPAs durch 4 unterschiedliche Supervidierende für eine reliable Bewertung notwendig sind. Die G-Koeffizienten  $>0,7$  entsprechen dabei den Empfehlungen für Assessments wie OSCEs oder andere hochwertige Assessments [26], [68]. Wir beobachteten sogar höhere phi-Werte für kleine Weiterbildungsstätten, was darauf hindeutet, dass kleinere zu größeren hier mindestens vergleichbar sind. Die in dieser Studie verwendeten App-basierte Assessments lieferten also zuverlässige Ergebnisse für Kombinationen von je 3 verwendeten 10 EPAs. Bei routinemäßiger Anwendung

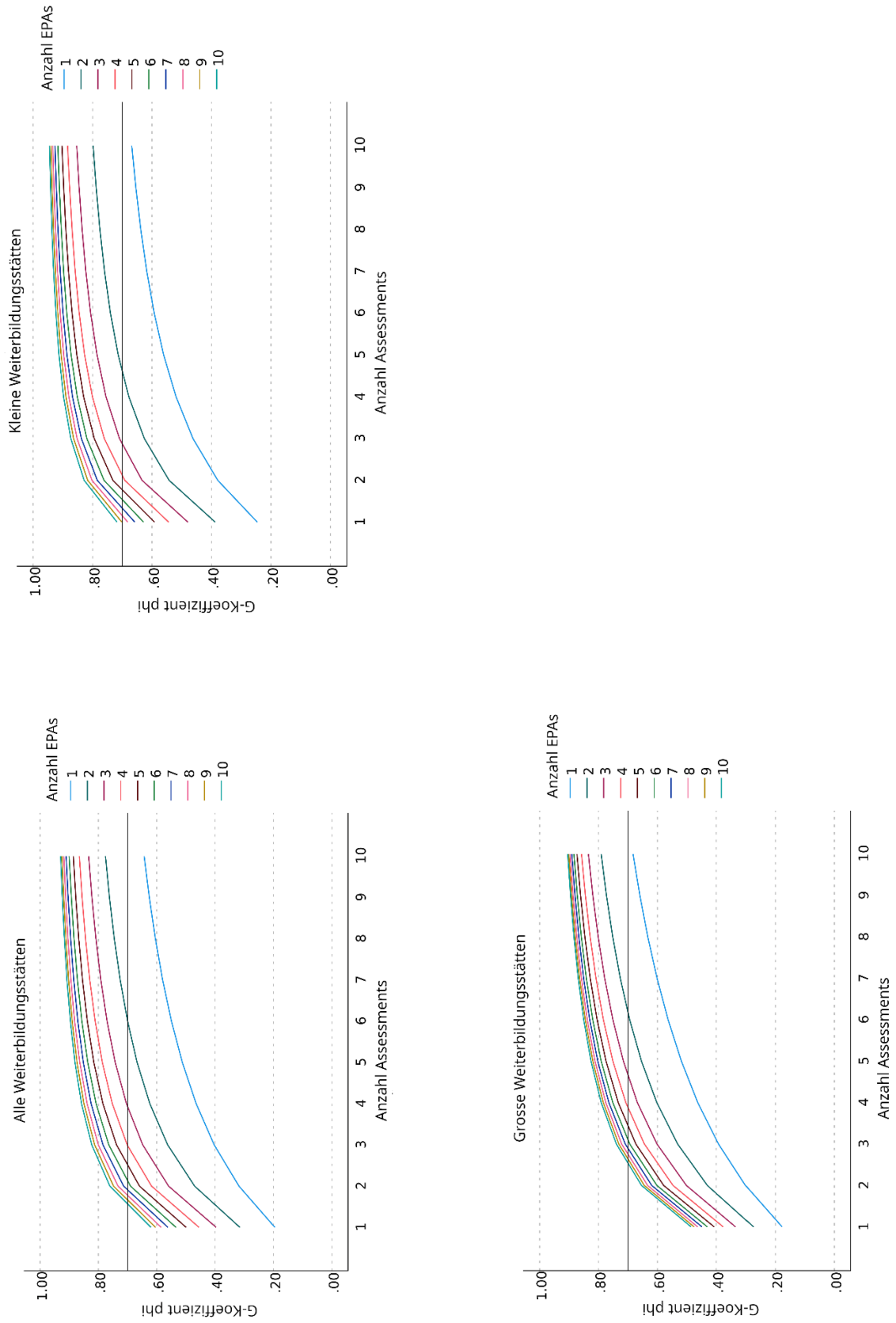


Abbildung 1: D-Studien zu Tabelle 4



ist es zu erwarten, dass die meisten Trainees diese Zahlen schnell erreichen, da sie bis zum Ende des zweiten Weiterbildungsjahres beherrscht werden sollten [56]. Ein typisches Portfolio wird also eine Vielzahl von Assessments enthalten und unterstützt die Entscheidung der Supervidierenden, welche klinischen Aufgaben den Weiterzubildenden anvertraut werden können. Es ist anzunehmen, dass häufige Beobachtungen und Feedback während der täglichen Routine die Weiterbildung verbessern, wie es auch für ein „programmatic assessment“ vorgesehen ist. Wir können nun belegen, dass die Empfehlung solcher häufigen Assessments unabhängig der Größe der Weiterbildungsstätte möglich ist.

#### 4.4. Limitationen

Eine mögliche Quelle für Verzerrungen ist die freiwillige Teilnahme sowohl der Supervidierenden als auch der Trainees, was zu einer nicht-repräsentativen Stichprobe für beide Gruppen führen könnte. Darüber hinaus bleibt offen, welche praktische Bedeutung diese Bewertung zufälliger Kombinationen von drei verschiedenen EPAs (aus 10) haben kann. Nichtsdestotrotz könnten diese Daten Erkenntnisse für die künftige Umsetzung liefern. Die Übertragbarkeit der Ergebnisse auf obligatorische Bedingungen zu untersuchen, ist entscheidend. Daten über potenzielle Einflussfaktoren wie Geschlecht, Alter, Ausbildungsjahre usw. waren aufgrund der strengen Datenschutzrichtlinien, wie bereits erwähnt, nicht verfügbar. Da diese Faktoren wahrscheinlich einen Einfluss auf die Bildungsprozesse haben, hätten die Ergebnisse anders ausfallen können, wenn sie um diese Faktoren korrigiert worden wären. In dieser Studie wurden jedoch Weiterbildungsstätten in zwei Gruppen verglichen, und die Verteilung der Störfaktoren dürfte ähnlich sein, zumal Rotationen in kleine und große Weiterbildungsstätten in der Schweizer Anästhesie-Weiterbildung obligatorisch sind.

### 5. Schlussfolgerung

Häufige WBAs von EPAs mit einer mobilen Smartphone-App erwiesen sich sowohl in großen als auch in kleinen Weiterbildungsstätten als zuverlässig. Wir fanden keine relevanten Unterschiede zwischen den Gruppen hinsichtlich der Anzahl und Verteilung der durchgeführten Assessments. Die Auswertung der G-Theorie und der D-Studien bestätigte, dass die Mindestanzahl an Assessments für einen g-Koeffizienten  $>0,7$  in beiden Gruppen erreicht werden kann. Daher scheint die Unterstützung der kompetenzbasierten Weiterbildung durch mobile Apps zur Bewertung von EPAs für große und kleine Anästhesie-Weiterbildungsstätten geeignet.

### ORCID der Autor\*innen

- Tobias Tessmann: [0000-0002-4687-7452]
- Adrian P. Marty: [0000-0003-3452-9730]
- Daniel Stricker: [0000-0002-7722-0293]
- Sören Huwendiek: [0000-0001-6116-9633]
- Jan Breckwoldt: [0000-0003-1716-1970]

### Interessenkonflikt

APM ist Mitglied der Weiterbildungskommission der „Swiss Society of Anesthesia and Perioperative Medicine“ (SSAPM) und Mitglied der EPA-Kommission des „Schweizerischen Instituts für ärztliche Weiter- und Fortbildung“ (SIWF). Mit Fördergeldern aus dem „Kompetitiven Lehrkredit“ der Universität Zürich und einer Projektförderung des SIWF wurde 2019 ein erster funktionaler Prototyp von einer externen Softwarefirma entwickelt. Im Herbst 2020 gründete APM eine Firma (precisionED AG), um die App von Grund auf neu zu entwickeln und ein nachhaltiges, qualitativ hochwertiges Assessment-System bereitzustellen. precisionED hält alle Rechte am geistigen Eigentum und garantiert einen zeitgemäßen Schutz der Daten unter Einhaltung der GDPR-Standards. APM ist Chief Visionary Officer und Verwaltungsratspräsident der precisionED AG.

SH ist Mitglied der EPA-Kommission des SIWF. JB ist Mitglied der EPA-Kommission und Co-Vorsitzender des „teach the teacher“ Programms der SIWF.

### Literatur

1. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver I, Taber S, Talbot M, Harris KA. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645. DOI: 10.3109/0142159X.2010.501190
2. Park YS, Hodges BD, Tekian A. Evaluating the Paradigm Shift from Time-Based Toward Competency-Based Medical Education: Implications for Curriculum and Assessment. In: Wimmers PF, Mentkowski M, editors. *Assessing Competence in Professional Performance across Disciplines and Professions*. Cham: Springer International Publishing; 2016. p.411-425. DOI: 10.1007/978-3-319-30064-1\_19
3. Jonker G, Hoff RG, Ten Cate OTJ. A case for competency-based anaesthesiology training with entrustable professional activities: an agenda for development and research. *Eur J Anaesthesiol*. 2015;32(2):71-76. DOI: 10.1097/EJA.000000000000109
4. Breckwoldt J, Beckers SK, Breuer G, Marty A. „Entrustable professional activities“: Zukunftsweisendes Konzept für die ärztliche Weiterbildung [Entrustable professional activities: Promising concept in postgraduate medical education]. *Anaesthesist*. 2018;67(6):452-457. DOI: 10.1007/s00101-018-0420-y
5. Englander R, Carraccio C. From theory to practice: making entrustable professional activities come to life in the context of milestones. *Acad Med*. 2014;89(10):1321-1323. DOI: 10.1097/ACM.0000000000000324

6. Ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ.* 2005;39(12):1176-1177. DOI: 10.1111/j.1365-2929.2005.02341.x
7. Weller JM, Misur M, Nicolson S, Morris J, Ure S, Crossley J, Jolly B. Can I leave the theatre? A key to more reliable workplace-based assessment. *Br J Anaesth.* 2014;112(6):1083-1091. DOI: 10.1093/bja/aeu052
8. Pinilla S, Lerch S, Lüdi R, Neubauer F, Feller S, Stricker D, Berendonk C, Huwendiek S. Entrustment versus performance scale in high-stakes OSCEs: Rater insights and psychometric properties. *Med Teach.* 2023;45(8):885-892. DOI: 10.1080/0142159X.2023.2187683
9. Norcini J, Anderson MB, Bolla V, Burch V, Costa MJ, Duvivier R, Hays R, Palacios Mackay MF, Roberts T, Swanson D. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-1109. DOI: 10.1080/0142159X.2018.1500016
10. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
11. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
12. Van Der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. Twelve Tips for programmatic assessment. *Med Teach.* 2015;37(7):641-646. DOI: 10.3109/0142159X.2014.973388
13. Rich JV, Fostaty Young S, Donnelly C, Hall AK, Dagnone JD, Weersink K, Caudle J, Van Melle E, Klinger DA. Competency-based education calls for programmatic assessment: But what does this look like in practice? *J Eval Clin Pract.* 2020;26(4):1087-1095. DOI: 10.1111/jep.13328
14. Heeneman S, de Jong LH, Dawson LJ, Wilkinson TJ, Ryan A, Tait GR, Rice N, Torre D, Freeman A, van der Vleuten CP. Ottawa 2020 consensus statement for programmatic assessment - 1. Agreement on the principles. *Med Teach.* 2021;43(10):1139-1148. DOI: 10.1080/0142159X.2021.1957088
15. Marty AP, Linsenmeyer M, George B, Young JQ, Breckwoldt J, ten Cate O. Mobile technologies to support workplace-based assessment for entrustment decisions: Guidelines for programs and educators: AMEE Guide No. 154. *Med Teach.* 2023;45(11):1203-1213. DOI: 10.1080/0142159X.2023.2168527
16. Young JQ, Sugarman R, Schwartz J, McClure M, O'Sullivan PS. A mobile app to capture EPA assessment data: Utilizing the consolidated framework for implementation research to identify enablers and barriers to engagement. *Perspect Med Educ.* 2020;9(4):210-219. DOI: 10.1007/s40037-020-00587-z
17. Young JQ, McClure M. Fast, Easy, and Good: Assessing Entrustable Professional Activities in Psychiatry Residents With a Mobile App. *Acad Med.* 2020;95(10):1546-1549. DOI: 10.1097/ACM.0000000000003390
18. George BC, Bohnen JD, Schuller MC, Fryer JP. Using smartphones for trainee performance assessment: A SIMPL case study. *Surgery.* 2020;167(6):903-906. DOI: 10.1016/j.surg.2019.09.011
19. Lefroy J, Roberts N, Molyneux A, Bartlett M, Gay S, McKinley R. Utility of an app-based system to improve feedback following workplace-based assessment. *Int J Med Educ.* 2017;8:207-216. DOI: 10.5116/ijme.5910.dc69
20. Findyartini A, Raharjanti NW, Greviana N, Prajogi GB, Setyorini D. Development of an app-based e-portfolio in postgraduate medical education using Entrustable Professional Activities (EPA) framework: Challenges in a resource-limited setting. *TAPS.* 2021;6(4):92-106. DOI: 10.29060/TAPS.2021-6-4/OA2459
21. Duggan N, Curran VR, Fairbridge NA, Deacon D, Coombs H, Stringer K, Pennell S. Using mobile technology in assessment of entrustable professional activities in undergraduate medical education. *Perspect Med Educ.* 2021;10(6):373-377. DOI: 10.1007/s40037-020-00618-9
22. Diwersi N, Gass JM, Fischer H, Metzger J, Knobe M, Marty AP. Surgery goes EPA (Entrustable Professional Activity) - how a strikingly easy to use app revolutionizes assessments of clinical skills in surgical training. *BMC Med Educ.* 2022;22(1):559. DOI: 10.1186/s12909-022-03622-1
23. Marty AP, Braun J, Schick C, Zalunardo MP, Spahn DR, Breckwoldt J. A mobile application to facilitate implementation of programmatic assessment in anaesthesia training. *Br J Anaesth.* 2022;128(6):990-996. DOI: 10.1016/j.bja.2022.02.038
24. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ.* 2011;45(6):560-569. DOI: 10.1111/j.1365-2923.2010.03913.x
25. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ.* 2012;46(1):28-37. DOI: 10.1111/j.1365-2923.2011.04166.x
26. Peeters MJ, Cor MK. Guidance for high-stakes testing within pharmacy educational assessment. *Curr Pharm Teach Learn.* 2020;12(1):1-4. DOI: 10.1016/j.cptl.2019.10.001
27. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach.* 2012;34(11):960-992. DOI: 10.3109/0142159X.2012.703791
28. Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach.* 2013;35(6):e1197-e1210. DOI: 10.3109/0142159X.2013.788789
29. Andersen SA, Nayahangan LJ, Park YS, Konge L. Use of Generalizability Theory for Exploring Reliability of and Sources of Variance in Assessment of Technical Skills: A Systematic Review and Meta-Analysis. *Acad Med.* 2021;96(11):1609-1619. DOI: 10.1097/ACM.0000000000004150
30. Kreiter C, Zaidi N. Generalizability Theory's Role in Validity Research: Innovative Applications in Health Science Education. *Health Prof Educ.* 2020;6(2):282-290. DOI: 10.1016/j.hpe.2020.02.002
31. Dunne D, Gielissen K, Slade M, Park YS, Green M. WBAs in UME- How Many Are Needed? A Reliability Analysis of 5 AAMC Core EPAs Implemented in the Internal Medicine Clerkship. *J Gen Intern Med.* 2022;37(11):2684-2690. DOI: 10.1007/s11606-021-07151-3
32. Kelleher M, Kinnear B, Sall D, Schumacher D, Schauer DP, Warm EJ, Kelcey B. A Reliability Analysis of Entrustment-Derived Workplace-Based Assessments. *Acad Med.* 2020;95(4):616-622. DOI: 10.1097/ACM.0000000000002997
33. Alves de Lima A, Conde D, Costabel J, Corso J, Van der Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. *Adv Health Sci Educ Theory Pract.* 2013;18(1):5-13. DOI: 10.1007/s10459-011-9343-y
34. Sullivan GM. A Primer on the Validity of Assessment Instruments. *J Grad Med Educ.* 2011;3(2):119-120. DOI: 10.4300/JGME-D-11-00075.1
35. Rourke J. WHO Recommendations to improve retention of rural and remote health workers - important for all countries. *Rural Remote Health.* 2010;10(4):1654. DOI: 10.22605/RRH1654

36. Somporn P, Ash J, Walters L. Stakeholder views of rural community-based medical education: a narrative review of the international literature. *Med Educ.* 2018;52(8):791-802. DOI: 10.1111/medu.13580
37. Health Education England. *Training in Smaller Places.* London: Health Education England; 2016.
38. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach.* 2001;23(5):445-454. DOI: 10.1080/01421590120075661
39. Boonluksiri P, Thongmak T, Warachit B. Comparison of educational environments in different sized rural hospitals during a longitudinal integrated clerkship in Thailand. *Rural Remote Health.* 2021;21(4):6883. DOI: 10.22605/RRH6883
40. Carmody DF, Jacques A, Denz-Penhey H, Puddey I, Newnham JP. Perceptions by medical students of their educational environment for obstetrics and gynaecology in metropolitan and rural teaching sites. *Med Teach.* 2009;31(12):e596-602. DOI: 10.3109/01421590903193596
41. Parry J, Mathers J, Al-Fares A, Mohammad M, Nandakumar M, Tsivos D. Hostile teaching hospitals and friendly district general hospitals: final year students' views on clinical attachment locations. *Med Educ.* 2002;36(12):1131-1141. DOI: 10.1046/j.1365-2923.2002.01374.x
42. Worley P, Esterman A, Prideaux D. Cohort study of examination performance of undergraduate medical students learning in community settings. *BMJ.* 2004;328(7433):207-209. DOI: 10.1136/bmj.328.7433.207
43. Lyon PM, McLean R, Hyde S, Hendry G. Students' perceptions of clinical attachments across rural and metropolitan settings. *Ass Eval High Educ.* 2008;33(1):63-73. DOI: 10.1080/02602930601122852
44. Kumar N, Brooke A. Should we teach and train in smaller hospitals? *Future Healthc J.* 2020;7(1):8-11. DOI: 10.7861/fhj.2019-0056
45. Bernabeo EC, Holtman MC, Ginsburg S, Rosenbaum JR, Holmboe ES. Lost in Transition: The Experience and Impact of Frequent Changes in the Inpatient Learning Environment. *Acad Med.* 2011;86(5):591-598. DOI: 10.1097/ACM.0b013e318212c2c9
46. Bonnie LHA, Cremers GR, Nasori M, Kramer AWM, van Dijk N. Longitudinal training models for entrusting students with independent patient care?: A systematic review. *Med Educ.* 2022;56(2):159-169. DOI: 10.1111/medu.14607
47. Simpson M, Tuck N, Bellamy S. Small business success factors: the role of education and training. *Educ Train.* 2004;46(8/9):481-491. DOI: 10.1108/00400910410569605
48. Anselmann S. Learning barriers at the workplace: Development and validation of a measurement instrument. *Front Educ.* 2022;7:880778. DOI: 10.3389/educ.2022.880778
49. Ashton D, Sung J, Raddon A, Riordan T. Challenging the myths about learning and training in small and medium-sized enterprises: Implications for public policy. *Employment Working Paper No. 1.* Geneva: International Labour Organization; 2008. p.1:65.
50. Panagiotakopoulos A. Barriers to employee training and learning in small and medium-sized enterprises (SMEs). *Develop Learn Organ.* 2011;25(3):15-18. DOI: 10.1108/14777281111125354
51. Jones P, Beynon MJ, Pickernell D, Packham G. Evaluating the Impact of Different Training Methods on SME Business Performance. *Environ Plann C Gov Policy.* 2013;31(1):56-81. DOI: 10.1068/c12113b
52. Attwell G, Deitmer L. Developing work based personal learning environments in small and medium enterprises. *PLE Conference Proceedings.* 2012;1(1). Zugänglich unter/available from: <https://proa.ua.pt/index.php/ple/article/view/16473>
53. Kock H, Ellström P. Formal and integrated strategies for competence development in SMEs. *J Eur Industr Train.* 2011;35(1):71-88. DOI: 10.1108/030905911111095745
54. Kravčík M, Neulinger K, Klamma R. Boosting Vocational Education and Training in Small Enterprises. In: Verbert K, Sharples M, Klobučar T, editors. *Adaptive and Adaptable Learning.* Cham: Springer International Publishing; 2016. p.600-604. DOI: 10.1007/978-3-319-45153-4\_72
55. Dawe S, Nguyen N. Education and Training that Meets the Needs of Small Business: A Systematic Review of Research. *National Centre for Vocational Education Research Ltd; 2007.* Zugänglich unter/available from: <https://eric.ed.gov/?id=ED499699>
56. Marty AP, Schmelzer S, Thomasin RA, Braun J, Zalunardo MP, Spahn DR, Breckwoldt J. Agreement between trainees and supervisors on first-year entrustable professional activities for anaesthesia training. *Br J Anaesth.* 2020;125(1):98-103. DOI: 10.1016/j.bja.2020.04.009
57. ten Cate O. Nuts and Bolts of Entrustable Professional Activities. *J Grad Med Educ.* 2013;5(1):157-158. DOI: 10.4300/JGME-D-12-00380.1
58. ten Cate O, Schwartz A, Chen HC. Assessing Trainees and Making Entrustment Decisions: On the Nature and Use of Entrustment-Supervision Scales. *Acad Med.* 2020;95(11):1662-1669. DOI: 10.1097/ACM.0000000000003427
59. ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach.* 2015;37(11):983-1002. DOI: 10.3109/0142159X.2015.1060308
60. Brennan R, Norman G. *G\_String, A Windows wrapper for urGENOVA®.* Hamilton: McMaster Education Research, Innovation & Theory (MERIT); 2017. Zugänglich unter/available from: [http://fhsperd.mcmaster.ca/g\\_string/index.html](http://fhsperd.mcmaster.ca/g_string/index.html)
61. Hartmann K. Assistentenstellen 2021 pro Fachgebiet, alle Weiterbildungsstätten. 6. Oktober 2021
62. Woodworth GE, Marty AP, Tanaka PP, Ambardekar AP, Chen F, Duncan MJ, Fromer IR, Hallman MR, Klesius LL, Ladlie BL, Mitchel SA, Miller Juve AK, McGrath BJ, Shepler JA, Sims 3rd C, Spofford CM, Van Cleve W, Maniker RB. Development and Pilot Testing of Entrustable Professional Activities for US Anesthesiology Residency Training. *Anesth Analg.* 2021;132(6):1579-1591. DOI: 10.1213/ANE.0000000000005434
63. Telio S, Ajjawi R, Regehr G. The "Educational Alliance" as a Framework for Reconceptualizing Feedback in Medical Education. *Acad Med.* 2015;90(5):609-614. DOI: 10.1097/ACM.0000000000000560
64. Brenner M, Weiss-Breckwoldt AN, Condrau F, Breckwoldt J. Does the 'Educational Alliance' conceptualize the student - supervisor relationship when conducting a master thesis in medicine? An interview study. *BMC Med Educ.* 2023;23(1):611. DOI: 10.1186/s12909-023-04593-7
65. Hauer KE, Ten Cate O, Boscardin C, Irby DM, Iobst W, O'Sullivan PS. Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ Theory Pract.* 2014;19(3):435-456. DOI: 10.1007/s10459-013-9474-4
66. Castanelli DJ, Weller JM, Molloy E, Bearman M. Trust, power and learning in workplace-based assessment: The trainee perspective. *Med Educ.* 2022;56(3):280-291. DOI: 10.1111/medu.14631

67. Gille F. About the Essence of Trust: Tell the Truth and Let Me Choose—I Might Trust You. *Int J Public Health*. 2022;67:1604592. DOI: 10.3389/ijph.2022.1604592
68. American Educational Research Association, editor. Standards for Educational & Psychological Testing. Washington, D.C: American Educational Research Association; 2014. p.60.

**Korrespondenzadresse:**

Prof. Dr. Jan Breckwoldt  
Universitätsspital Zürich, Institut für Anästhesiologie,  
Raemistr. 100, CH-8091 Zürich, Schweiz  
jan.breckwoldt@usz.ch

**Bitte zitieren als**

Tessmann T, Marty AP, Stricker D, Huwendiek S, Breckwoldt J. There is no "too small" for frequent workplace-based assessment: Differences between large and small residency programs in anesthesia when using a mobile application to assess EPAs. *GMS J Med Educ*. 2024;41(5):Doc54. DOI: 10.3205/zma001709, URN: urn:nbn:de:0183-zma0017091

**Artikel online frei zugänglich unter**

<https://doi.org/10.3205/zma001709>

**Eingereicht:** 19.10.2023

**Überarbeitet:** 08.05.2024

**Angenommen:** 11.06.2024

**Veröffentlicht:** 15.11.2024

**Copyright**

©2024 Tessmann et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.