Peak and ceiling effects in final-product analysis of mastoidectomy performance

Niels West, MD[1], Lars Konge, MD, PhD[2], Per Cayé-Thomasen, MD, DMSc[1], Mads Sølvsten Sørensen, MD, DMSc[1], Steven Arild Wuyts Andersen, MD[1]

1. Department of Otorhinolaryngology – Head & Neck Surgery, Rigshospitalet, Copenhagen, Denmark.
2. Center for Clinical Education (CEKU), Centre for HR, The Capital Region, Copenhagen, Denmark.

Corresponding author: Steven A. W. Andersen, MD, Department of Otorhinolaryngology – Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark. E-mail: stevenarild@gmail.com. Phone: 0045 20612006.

Previous presentations: None.

Full citation: West N, Konge L, Cayé-Thomasen P, Sørensen MS, Andersen SA. Peak and ceiling effects in final-product analysis of mastoidectomy performance. J Laryngol Otol. 2015 Nov; 129(11):1091–6.

DOI: 10.1017/S0022215115002364

Introduction: Virtual reality surgical simulation of mastoidectomy is a promising training tool for novices. Final-product analysis for assessment of novice mastoidectomy performance could be limited by a peak or ceiling effect. These could possibly be countered by simulator-integrated tutoring.

Methods: Twenty-two participants completed a single session of self-directed practice of the mastoidectomy procedure in a VR simulator. Participants were randomized for additional simulator-integrated tutoring. Performances were assessed at 10-minute intervals using final-product analysis.

Results: 45.5% of participants peaked before the 60-minute time limit. None of the participants achieved the maximum score suggesting a ceiling effect. The tutored group performed better than the non-tutored group but tutoring did not eliminate the peak or ceiling effects.

Conclusion: Timing and adequate instruction is important when using final-product analysis for the assessment of novice mastoidectomy performance. Improved real-time feedback and tutoring could address the limitations of final-product-based assessment.

Key words: Temporal Bone; Education, Medical; Computer Simulation; Clinical competence

Introduction

Temporal bone surgery is one of the key skills for the otorhinolaryngology resident. A shortage of cadaveric temporal bones makes virtual reality (VR) simulation training a promising alternative to traditional cadaveric dissection training of mastoidectomy[1,2]. In addition, self-directed training using simulator-integrated tutoring could potentially reduce the need for dedicated expert instructors, who are often limited by busy schedules and clinical duties. Residents could acquire some of the basic competencies in the virtual environment, saving cadavers and costly instructional resources for subsequent and more advanced training. Tutoring in medical technical skills training is diverse and has been demonstrated to be effective in different settings[3-5]. The role of simulator-integrated tutoring in facilitating mastoidectomy skills remains largely unexplored[6].

Assessment of performance is essential to monitor and track the progress of the trainee's technical skills and to establish the effect of learning interventions such as VR simulation training. A number of assessment tools for evaluating mastoidectomy performance have been developed[7-9]; these are based either on the process, such as global rating scales[10], requiring live or recorded observation of the performance, or on rating the final-product[11]. Final-product analysis (FPA) does not consider how the goal is reached and is normally used for performance assessment after the procedure is finished. Nevertheless, final-product-like items are found to be relevant in the assessment of mastoidectomy competency[12,13]. FPA could be hypothesized to have limitations if used to monitor progress and technical skills development in novices; if FPA is not applied at the optimal point in time, the FPA-score will not necessary reflect the peak performance but instead a lower performance—either because the novice was not allowed adequate time for the procedure or because the novice lacks knowledge on when to stop and proceeds to damage vital structures. In addition to this peak effect, FPA could be limited by a ceiling effect in which a plateau in performance is reached because the scale lacks discriminate power[14,15] or simply because self-directed training alone is not sufficient for novices to progress beyond a certain level.

In this study, we aim to establish whether ceiling and peak effects limit FPA of novice mastoidectomy performance in a VR temporal bone simulator. We also want to investigate if additional simulator-integrated tutoring can counter these effects.

Materials and methods

Twenty-two medical students from the University of Copenhagen, Denmark, volunteered for participation and signed informed consent. They were all novices with no previous exposure to or experience with temporal bone surgery. All participants received a one-hour introductory lecture on temporal bone anatomy, the mastoidectomy procedure, and the VR simulator.

The participants were then asked to perform a complete mastoidectomy with posterior tympanotomy in the Visible Ear Simulator (VES). The VES is a freeware real-time 3D virtual temporal bone simulator that can be downloaded from the Internet[16]. The software runs on a standard PC with a Nvidia Geforce GTX™ graphics card and supports the Geomagic Touch™ (3D Systems, USA) haptic device for force-feedback and intuitive drilling[17,18]. VES provides a step-by-step tutorial of the surgical procedure with text and illustrations from the simulator and the option of enabling an accompanying integrated tutor-function that green-lights the volume to be drilled in each step of the tutorial (Figure 1). An experimental version of the simulator was developed in order to auto-save successive copies of the virtual temporal bones at pre-defined intervals during the procedure.

The participants were allowed 60 minutes for the procedure. All participants were self-directed and provided with the on-screen step-by-step tutorial to the procedure. Participants were randomized to have the on-screen tutorial supplemented by the simulator-integrated tutoring (green-lighting). Randomization was performed as quasi-randomization to the day of the training using computer-based assignment.

The simulator auto-saved the virtual temporal bones at 10-minute intervals between 20 and 60 minutes. The final-products were later analysed by two senior otologists blinded to the participant, the time of save and the use of tutoring. The final-product saved-files were opened in the simulator and rated using a modified Welling Scale with 26 binary items for assessment of mastoidectomy performance[12].

Data were analysed with SPSS (SPSS Inc., Chicago, IL) version 22 for MacOS X using parametric statistics as data were normally distributed according to the Shapiro-Wilks test. In comparing continuous data, dependent and independent sample t-tests were performed as appropriate. Fisher’s exact test was used for comparing categorical data. Results were considered significant if $p<0.05$.

Results and analysis

Baseline characteristics regarding age, sex, years of medical studies and simulation and computer-related experience and interest are found in Table I. Baseline characteristics were without statistically significant differences between the tutored and the non-tutored groups.

The mean final-product score of all participants was 17.3 at 60 minutes (the maximum allowed time) (Table II). Consistent with a peak effect, the highest score at any time-point during the procedure—the peak score—was found to be higher (18.0, $p<0.01$). None of the participants achieved the maximum score of 26 and only 3 of the participants were able to reach final-product scores of more than 20 at some point during the procedure reflecting a ceiling effect.

For further analysis, we divided the participants according to whether their scores peaked during the procedure ('early peakers') or not ('late peakers')(Table II). Ten of the 22 participants (45.5 %) achieved their highest score (peak score) before the full, allowed 60 minutes and their final-product performance progressively deteriorated after peaking early. The mean peak score was significantly higher in the 'early peakers' group compared with the 'late peakers' group (18.9 vs. 17.2, $p<0.01$)(Table II). As a group the 'early peakers' consistently outperformed the 'late peakers' during the procedure except at 60 minutes where both groups ended up having equal scores (Figure 2).

Participants randomized to simulator-integrated tutoring with green-lighting had a performance similar to the non-tutored participants during the first 40 minutes after which the tutored participants outperformed non-tutored participants (Figure 3). This was reflected in significantly higher mean peak and 60-minute scores for the tutored group (Table II). Even though simulator-integrated tutoring led to a better performance, it did not eliminate the peak and ceiling effects; a similar distribution of tutored and non-tutored participants was found in the 'early peakers' and 'late peakers' groups using Fisher's exact test ($p=0.69$) and the 15% highest scores were achieved by tutored and non-tutored participants with a 2:1 distribution and the 25% highest scores with a 1:1 distribution.

Discussion

In this prospective study on the limitations of final-product analysis of mastoidectomy performance, we used a VR simulator with 10-minute interval auto-save to map final-product progress and found that peak and ceiling effects were at play. Additional simulator-integrated tutoring with green-lighting was not found to eliminate these effects.

There is little knowledge on peak and ceiling effects in final-product performance assessment. In most studies on mastoidectomy, novices are given a certain timeframe within to finish the procedure and FPA is applied to assess performance after the given time. This timeframe is assumed to be sufficient to finish the procedure for all novices and not continuing to do damage to the final product. However, this requires novices to know the procedure in depth beforehand. Little consideration has been given to the fact that not all novices possess this knowledge and might use any additional time to explore the temporal bone further. This behaviour would be natural and acceptable in a learning context but can be problematic in an assessment situation.

Our study enabled us to monitor progress by FPA at several time-points. In the given context—studying the performance of complete novices in self-directed learning—we found a peak effect: almost half of the participants had a peak final-product score before the end of the given timeframe and started making mistakes, thereby decreasing their score. This is exemplified by one participant who at 40 minutes achieved a score of 20 out of the maximum 26 points but finished with a score of 16 at the end of the session—the time point at which the final-product assessment would usually be employed. Whether the remaining participants that we designated 'late peakers' would have peaked at some point or plateaued if given more time is speculation. Nonetheless, time and performance are dependent factors and this should be considered when using FPA in mastoidectomy performance assessment.

The 'early peakers' demonstrate that novices can lack the knowledge of when to stop; in our study participants were self-directed and had only access to the on-screen step-by-step tutorial and for the tutored group additional green-lighting of the volume to be drilled in each step. In contrast to this, it has been established that feedback and directed goal setting are key elements in surgical technical skills acquisition[19-22]: in a study on VR thoracoscopy simulation training, an educator-guided group was found to perform significantly better than a self-directed group[21]; similar results have been demonstrated in VR colonoscopy simulation training[22]. In line with this, the theory of directed self-regulated learning (DSRL) includes an activated and mentally involved trainee with limited autonomous control in a structured setting, facilitating long-term learning[22-24].

In the literature on technical skills assessment, a ceiling effect most often concerns the limitation of an assessment tool in distinguishing between novices and experts. The blocking of further progress is another type of ceiling effect, relating to the simulation design and degree of difficulty[19,25]. The latter was found in our study; none of the participants achieved the maximum score and only a few participants achieved above 20 points. Whether this reflects a true ceiling or could be addressed solely by repeat practice warrants further investigation. A contributing factor to the ceiling effect could also be related to simulator fidelity and design. Nonetheless, the relative low performance in this self-directed setting and the literature on DSRL suggests that improvement in performance beyond this ceiling requires additional instruction, guidance and feedback; further analyses on where novices fail could contribute to qualify future instructions and simulator-integrated tutoring.

The additional simulator-integrated tutoring with volumetric green-lighting of the volume to be removed in each of the procedural steps significantly increased peak and final performance. The tutored group performed better than the non-tutored group especially in the later and more difficult parts of the procedure. However, simulator-integrated tutoring can be a double-edged sword: on one hand “perfect is the enemy of good”, meaning that the tutor could lead to or force a risky behaviour e.g. “expose more of the facial nerve” resulting in a reduced final-product performance; on the other hand, in a learning situation, this behaviour could facilitate acquisition of skills because novices are encouraged to perform the difficult tasks and can learn from their mistakes. Whereas simulator-integrated tutoring had a significant effect on performance it did no eliminate the peak or ceiling effect.

Our findings are consistent with the literature underpinning the limitations of self-directed VR simulation training of novices. For novices to acquire and improve surgical technical skills such as demanded by the mastoidectomy procedure there is a need for more than simply the VR simulation equipment. Additional instructional approaches and training are needed to fill the gap between the level of competency that can gained by VR simulation training alone and the appropriate level of competency before proceeding to the OR. Final-product-based assessment can be used to monitor the trainee’s progress and has a role if the limitations of FPA can be adequately addressed. Future automated simulator-based assessment is promising because multiple combined assessments could be feasible[13,26]. In combination with other simulator-gathered metrics, continuous FPA could be used to monitor progress in real-time during virtual practice and provide on-going feedback, guiding the novice to optimal performance. Individualized feedback and simulator-integrated tutoring could potentially enhance learning in self-directed training but a clearer definition of goals and progress monitoring is needed. Limited resources in surgical training could motivate these developments.

Conclusion

Timing of assessment and adequate instruction of trainees is important when using FPA for the assessment of novice mastoidectomy performance. Peak and ceiling effects could otherwise limit the final-product score and this should be considered even in the conventional application of final-product analysis. The current simulator-integrated tutor-function did not eliminate these effects, but future developments and improvements of on-going feedback and directed goal setting could address the limitations of final-product analysis.

Acknowledgements

The authors would like to acknowledge computer-graphics engineer Peter Trier Mikkelsen, MSc, for developing the Visible Ear Simulator including the experimental version used in this study. The authors would also like to thank Anders Kverneland, MD, for statistical support.

Financial Support: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflicts of Interests: None.

Ethical Standards: The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guidelines on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The regional ethics committee found that this study was exempt (H-4-2013-FSP-088).

References

1. George AP, De R. Review of temporal bone dissection teaching: how it was, is and will be. *J Laryngol Otol* 2010;**124**:119–25

2. Mills R, Lee P. Surgical skills training in middle-ear surgery. *J Laryngol Otol* 2003;**117**:159–63

3. Wiet GJ, Stredney D, Sessanna D, Bryan JA, Welling DB, Schmalbrock P. Virtual temporal bone dissection: an interactive surgical simulator. *Otolaryngol Head Neck Surg* 2002;**127**:79–83

4. Rhienmora P, Haddawy P, Khanal P, Suebnukarn S, Dailey MN. A virtual reality simulator for teaching and evaluating dental procedures. *Methods Inf Med* 2010;**49**:396–405

5. Ungi T, Sargent D, Moult E, Lasso A, Pinter C, McGraw RC *et al*. Perk Tutor: an open-source training platform for ultrasound-guided needle insertions. *IEEE Trans Biomed Eng* 2012;**59**:3475–81

6. Sewell C, Morris D, Blevins N, Dutta S, Agrawal S, Barbagli F *et al*. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg* 2008;**13**:63–81

7. Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective assessment of temporal bone drilling skills. *Ann Otol Rhinol Laryngol* 2007;**116**:793–8

8. Laeeq K, Bhatti NI, Carey JP, Della Santina CC, Limb CJ, Niparko JK *et al*. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope* 2009;**119**:2402–10

9. Butler NN, Wiet GJ. Reliability of the Welling scale (WS1) for rating temporal bone dissection performance. *Laryngoscope* 2007;**117**:1803–8

10. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchinson C *et al*. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;**84**:273–8

11. Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. *Am J Surg* 2000;**180**:234–7

12. Andersen SA, Cayé-Thomasen P, Sølvsten Sørensen M. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. *Laryngoscope* 2015;**125**:431–5

13. Wan D, Wiet GJ, Welling DB, Kerwin T, Stredney D. Creating a cross-institutional grading scale for temporal bone dissection. *Laryngoscope* 2010;**120**:1422–7

14. Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA. Ceiling effect in technical skills of surgical residents. *Am J Surg* 2004;**188**:294–300

15. Moorthy K, Munz Y, Adams S, Pandey V, Darzi A. A Human Factors Analysis of Technical and Team Skills Among Surgical Trainees During Procedural Simulations in a Simulated Operating Theatre. *Ann Surg* 2005;**242**:631–9

16. http://ves.cg.alexandra.dk/. Accessed on 15 December 2014

17. Sorensen MS, Mosegaard J, Trier P. The visible ear simulator: a public PC application for GPU-accelerated haptic 3D simulation of ear surgery based on the visible ear data. *Otol Neurotol* 2009;**30**:484–7

18. Trier P, Noe KØ, Sørensen MS, Mosegaard J. The visible ear surgery simulator. *Stud Health Technol Inform* 2008;**132**:523–5

19. Bech B, Lönn L, Falkenberg M, Bartholdy NJ, Räder SB, Schroeder TV *et al.* Construct validity and reliability of structured assessment of endoVascular expertise in a simulated setting. *Eur J Vasc Endovasc Surg* 2011;**42**:539–48

20. Reznick RK. Teaching and testing technical skills. *Am J Surg* 1993;**165**:358–361

21. Bjurström JM, Konge L, Lehnert P, Krogh CL, Hansen HJ, Petersen RH *et al.* Simulation-based training for thoracoscopy. *Sim Healthcare* 2013*:***8**:317–323

22. Kruglikova I, Grantcharov TP, Drewes AM, Funch-Jensen P. The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity Virtual-Reality simulation: a randomised controlled trial. *Gut* 2010:**59**:181–5

23. Brydges R, Carnahan H, Safir O, Dubrowski A. How effective is self-guided learning of clinical technical skills? It's all about process. *Med Educ* 2009:**43**:507–515

24. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. *Med Educ* 2012:**46**:648–656

25. Fann JI, Caffarelli AD, Georgette G, Howard SK, Gaba DM, Youngblood P *et al.* Improvement in coronary anastomosis with cardiac surgery simulation. *J Thorac Cardiovasc Surg* 2008;**136**:1486–91

26. Kerwin T, Wiet G, Stredney D, Shen H-W. Automatic scoring of virtual mastoidectomies using expert examples. *Int J Comput Assist Radiol Surg* 2012;**7**:1–11

Summary

- Final-product analysis (FPA) conventionally assesses performance based on the final-product at the end of the procedure but is in this study investigated as a progress-monitoring tool
- Peak and ceiling effects are possible limitations of FPA, but if adequately addressed FPA can potentially be integrated as an advanced feedback tool
- Timing of assessment and clear instruction of trainees is important when using FPA for the assessment of novice mastoidectomy performance

TABLE I
BASELINE CHARACTERISTICS

	All participants n=22	**Tutored participants** n=10	**Non-tutored participants** n=12
Age, average, years	25.3	24.9	25.7
Sex, female, n	8	5	3
Sex, male, n	14	5	9
Years of study, average	4.6	4.4	4.8
Any previous virtual surgical experience	55%	40%	67%
Computer usage, average/week, hours	17.0	18.7	15.5
Computer interest, average*	4.7	4.7	4.6
Self-rated IT-skills, average*	4.5	4.6	4.5
Previous gaming experience, average*	3.8	3.5	4.1

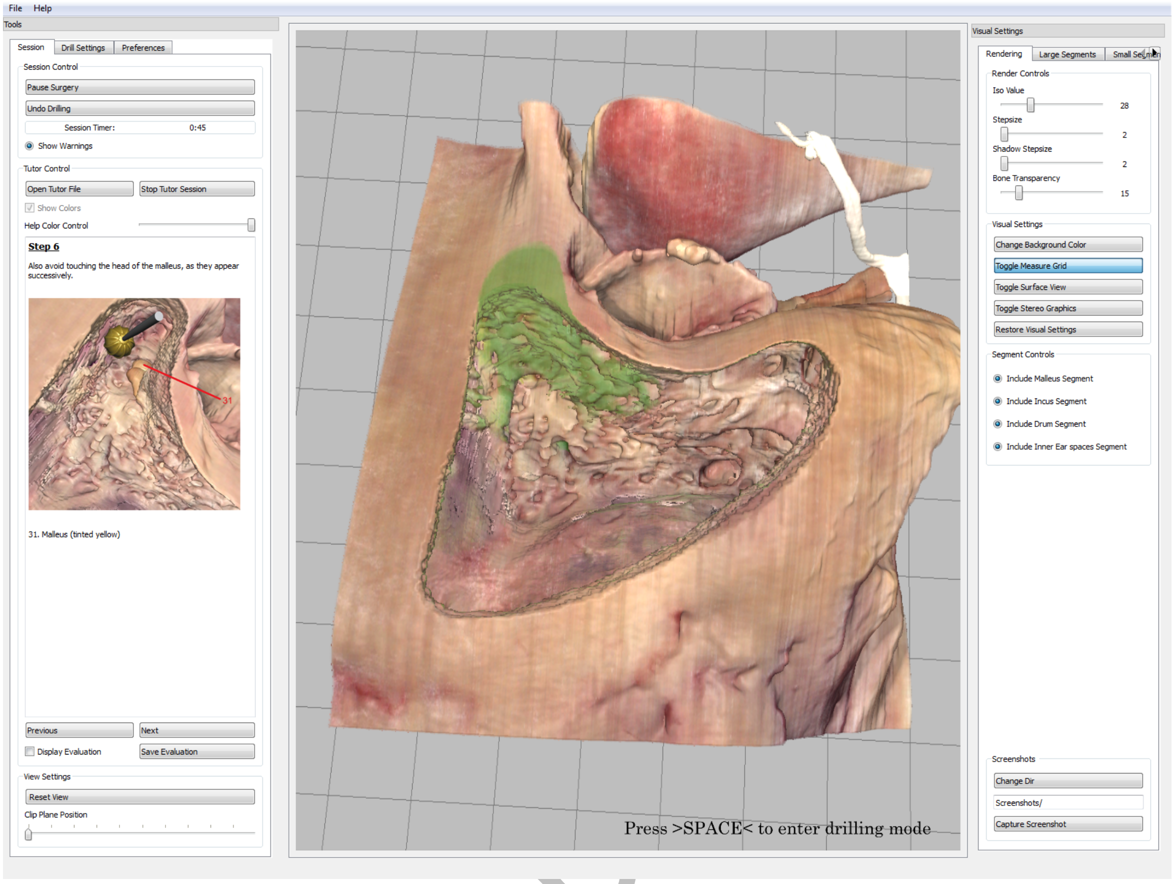
*On a 5 item Likert-like scale

TABLE II

MEAN FINAL-PRODUCT SCORES

	n	Peak score (95 % CI)	p	60-minute score (95 % CI)	p
All participants	22	18.0 (16.8–19.1)		17.3 (16.2–18.4)	
‘Early peakers’	10	18.9 (17.3–20.5)*	0.16	17.2 (15.7–18.6)*	0.96
‘Late peakers’	12	17.2 (15.6–18.8)		17.2 (15.6–18.8)	
Tutored participants	10	19.5 (18.4–20.6)	0.02	18.7 (17.4–19.9)	0.01
Non-tutored participants	12	16.7 (15.1–18.3)		16.0 (14.6–17.3)	

*The difference between the early peakers’ peak score and 60-minute score and was statistically significant ($p<0.01$). CI = Confidence Interval.



Session Timer: 0:45

● Show Warnings

☑ Show Colors

Step 6

Also avoid touching the head of the malleus, as they appear successively.

31. Malleus (tinted yellow)

□ Display Evaluation

Iso Value 28

Stepsize 2

Shadow Stepsize 2

Bone Transparency 15

● Include Malleus Segment

● Include Incus Segment

● Include Drum Segment

● Include Inner Ear spaces Segment

Screenshots/

Press >SPACE< to enter drilling mode

Fig.1 Screenshot from the simulator with the on-screen tutorial (to the left) and the corresponding green-lighted volume to be drilled in the step.

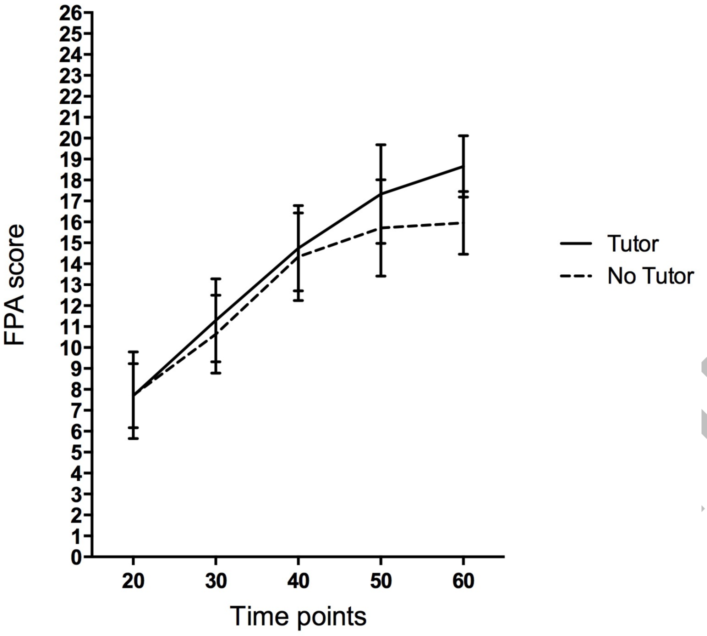


Fig.2. Performance for the tutored and non-tutored groups.

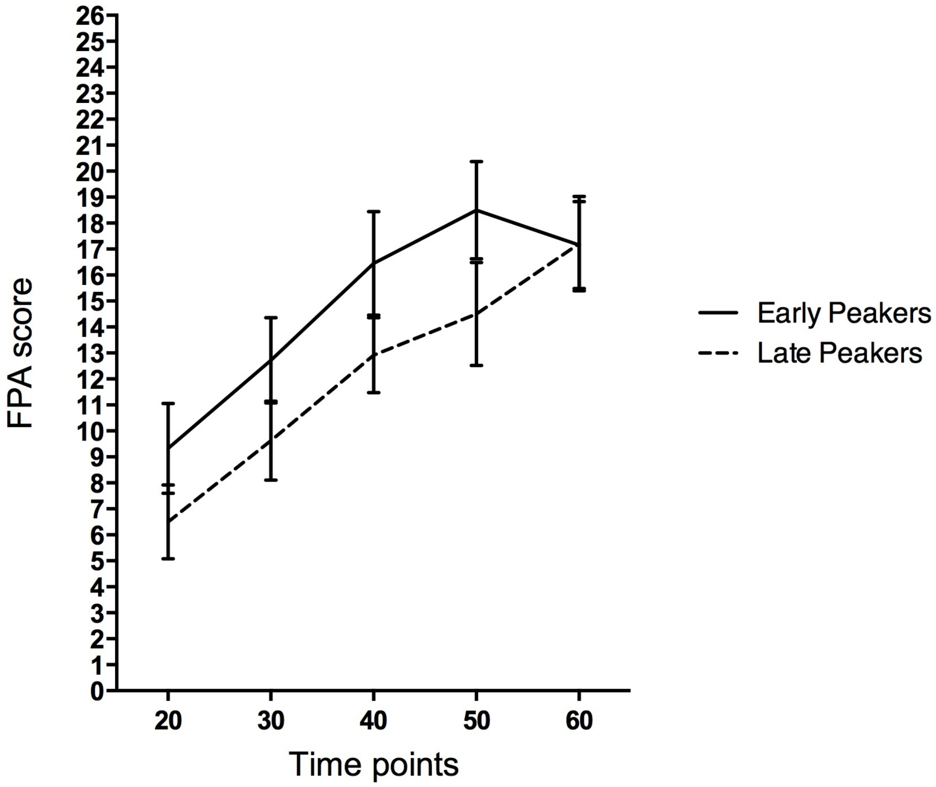


Fig.3. Performance of ‘early’ and ‘late’ peakers.