

RE: Lynch, J. and Porcellato, A. (2020) *The Case for an Aviation English Screening Tool for US Flight Schools* The Specialist 41, 4

4th November 2020

Dear Ms Lynch,

Dear Dr Porcellato,

We enjoyed reading your paper, *The Case for an Aviation English Screening Tool for US Flight Schools*, in the Specialist. Not only is it encouraging to see more focus on this area which has been given scant attention over the years, but we were interested to see Checkpoint discussed for the first time. It is gratifying that you recognise Checkpoint's strengths, as well as setting out areas which you believe could be improved.

We were surprised and disappointed that you didn't contact us as you were preparing your draft. An exchange prior to publication would have allowed us to share more information about Checkpoint which is not in the public domain, for example, how aspiring ATCs are addressed in test content, test taker familiarisation videos and our current development activities and research agenda, all of which would have improved the depth and accuracy of your discussion. In any case, we welcome the exposure and the critique – thank you!

Much of your paper echoes arguments that we have been making for some years, notably, the inappropriacy of screening aspiring pilots with tests of English for academic purposes and tests designed in accordance with the ICAO LPRs. We particularly liked your framing of the paper on student 'x'. Our mission in the domain of ab-initio aviation training aligns with yours - to help aspiring pilots and their sponsors and training organisations avoid problems associated with language proficiency, to help ensure that failure is avoided, that dreams are fulfilled.

Revision of the Checkpoint speaking assessment

The development of Checkpoint in 2013/14 was driven in part by a requirement from a European FTO for a screening tool for applicants to English-Medium Instruction (EMI) European Aviation Safety Agency (EASA) integrated Airline Transport Pilot Licence (ATPL) training. Such programmes (shown in purple in figure 1) differ considerably from FAA flight training, notably, that students spend a minimum of 750 hours (around 6 months) in the classroom undergoing Theoretical Knowledge (TK) training and examinations before they even see the inside of an aircraft cockpit.

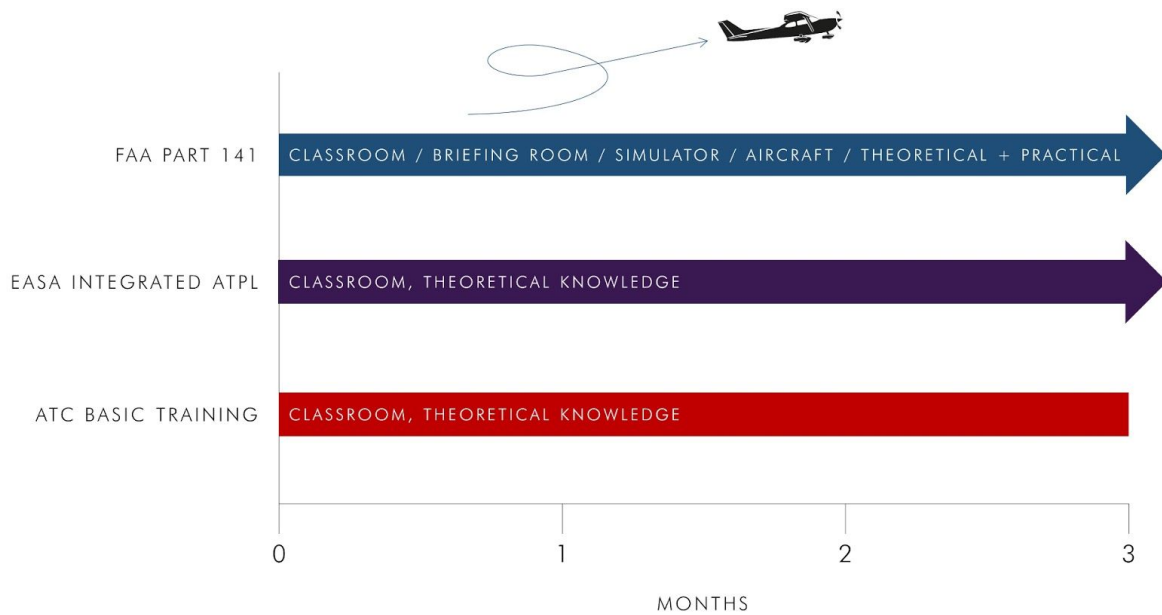


Figure 1: Comparison of the initial three months of training in various primary aviation training programmes

Under EASA ATPL training, speaking skills are necessary for effective transactional language use in an EMI environment. However, given the exclusive emphasis on instructor-led TK classroom instruction in the initial stages of training, coping successfully relies far more on receptive language skills than on spoken interactions (indeed, research suggests that students who score well on Checkpoint generally perform successfully in examinations and beyond in EASA ATPL training). Clearly, in FAA programmes (in blue in figure 1), students are involved in practical flying training so much earlier in their course. Given the context in which you work, we understand why you identify the need for deeper assessment of oral language proficiency, particularly at the higher end of spoken language abilities.

Today, Checkpoint is becoming more widely used by universities and flight schools that offer FAA flight training both in the USA and worldwide (the title of your paper is US-centric yet the subject matter is applicable not just to US flight schools but to the many FTOs around the world who offer EMI FAA part 141 training). Our stakeholders report that Checkpoint is both easy to use and provides scores which enable confident admissions and training decisions. Nevertheless, as Checkpoint was originally developed for European users, this year we began planning a revision of the Checkpoint speaking assessment to more closely address the needs of FTOs operating under the FAA flight training model. To support our own needs analysis, we wait with keen interest for the publication of Udell, Schneider and Kim's work. When complete, the revision will enhance the depth of measure of spoken

language proficiency for all Checkpoint users, and will tie in with further postgraduate investigations of test quality which we hope to pursue in 2021.

Screening for ab-initio pilots and ATCs

As figure 1 shows, basic ATC training in Europe and beyond shares much in common with the initial stages of EASA flight training. The EUROCONTROL Common Core Content for the initial 12 weeks of TK instruction contains content across similar subjects to flight training – aircraft performance, navigation, meteorology, etc. – all of which are delivered in a classroom with an instructor with supporting written courseware. In terms of screening for language proficiency, we would argue that for student pilots and ATCs in Europe, there are far more similarities than there are differences in the language proficiencies required for success in the early stages of training.

To address the needs of aspiring controllers at the point of selection and admission, Checkpoint already contains versions of the listening and speaking assessments with content oriented specifically to basic ATC training. This information is not on our website, an omission we shall address in the future.

We agree wholeheartedly with ICAEA that in the context of radiotelephony communications, licensed pilots and air traffic controllers have different language needs and therefore require different language test instruments. This guidance is unassailable. However, we question your assumption that the language needs of aspiring pilots and ATCs (particularly in Europe) are substantially different such that they warrant completely different screening instruments. In the absence of supporting evidence or, at the very least, principled discussion of the domain, we argue that there is no logic at all in applying industry guidance on testing for professional licensure to screening for successful primary instruction. In our field over the years, we have seen many examples of misinterpretation of industry guidance and arguably your point here falls into the same trap.

Context and practicality

Your paper's important but narrow focus on language content and tasks does not take into consideration the wider context of student selection and admission for flight training. To open the doors for students to pursue their dreams, FTOs need to assess a range of constructs of which language is just one. Others may include:

- Knowledge of maths and physics
- Hand-eye coordination and dexterity

- Cognitive reasoning
- Multi-tasking
- Personality; and
- Ability to pay

In our experience, pilot assessment varies on a continuum from a short telephone conversation to comprehensive 2-day assessment procedures involving a battery of tests, interviews and group interaction tasks. Language is either assessed implicitly, tangled up in assessments designed to tap non-language constructs, or explicitly in separate language proficiency tests such as Checkpoint. Given the breadth of measures, the time that stakeholders are able to give to language assessment is limited. Sometimes, it is none at all.

Your paper rightly advocates specific purpose language assessment which means abandoning widely available proficiency tests such as IELTS or TOEFL. However, it does not give consideration to the challenges that this raises: how to make assessment accessible to individuals who are often scattered around the world in widely disparate locations on one hand, and the limited time and resources that FTOs have available for language proficiency assessment on the other. Allied with this, your paper advocates building on the Checkpoint model by both replacing and adding tasks to assessments in all three skills. On the basis of domain and construct representation, we don't disagree with your position, but testing is inevitably a compromise between validity and practicality. Incorporating several substantial tasks across the skills would result in a screening test considerably longer than Checkpoint's 90 minutes (which, for many FTOs, is already prohibitively long). So, the question is as much 'what do we leave out?' as it is 'what do we include?' We would never suggest that practicality trumps adequate construct representation, but on first sight, resolving your wish list with the limited time and resources that stakeholders have for assessment would seem to be a significant challenge.

Language, tasks and subject matter knowledge

With regards to domain and construct representation, the categories that you use to distinguish between a flight training candidate and flight student are pertinent. Having made this important distinction, we should be especially careful not to blur the line beyond which a candidate passes from one to the other. Accordingly, whilst tasks such as listening to an ATIS, following ATC instructions, performing call-outs and reading NOTAMS are identified by Udell et al. as being fundamental to the flight training context, they are undoubtedly performed after some form of professional instruction or study. It is only then that such tasks

can be meaningful to test takers. Therefore we would strongly recommend avoiding them in a screening test. To do so would be to run the risk of:

1. Conflating language knowledge with subject matter knowledge with the attendant implications for construct-irrelevant variance in test scores, dangers which can be especially acute at lower levels of language proficiency where test takers may draw on subject matter knowledge in a compensatory way¹.
2. Conferring advantage on takers with subject matter knowledge and, vice versa, disadvantaging those without, quite possibly leading to negative washback.

It is important to remember that while needs analyses such as that conducted by Udell et. al. can greatly inform our work, they are not, in themselves, test specifications.

Face validity is a reasonable consideration in the construction of any assessment tool. Nonetheless, how a test looks is, for some, too artificial to be viewed as what your paper refers to as a critical factor in test design². In the construction of tasks, the question we must ask ourselves is 'are we assessing the ability to perform tasks such as understanding ATIS or checklist call-outs, or are we assessing the language proficiencies required to *learn how* to perform these tasks? Consequently, we would argue that construct validity should be our primary consideration in the construction of screening tests for aspiring aviation professionals.

Our position at Latitude is one of transparency. We hope this is evident from the information we present on Checkpoint in the public domain. While we are pleased to engage in debate, we would much prefer to work with our peers in the spirit of cooperation, especially given that the needs of learners are so great and our community of aviation English practitioners is so small. Should you wish to learn more about Checkpoint, or, even better, work with us in the area of assessment for student screening, please let us know. We would welcome discussion.

Best wishes,

The Latitude Team

¹ See, for example, Clapham, C. (1998) *The Effect of Language Proficiency and Background Knowledge on EAP Students' Reading Comprehension* in Kunnan, A. (ed.) *Validation in Language Assessment* Routledge, New York.

² See Bachman, L. (1990) *Fundamental considerations in Language Testing*, New York: OUP, pp 285- 289.