On the transportability of a computerised test battery for the selection of pilots

Agnes Kokorian
People Technologies
123 Molesey Park Road
East Molesey, Surrey KT8 0JX, UK
agneskokorian@ntlworld.com

1st. Lt. Psychologist Mauricio Pereira da Costa
Academia da Forca Aerea
Estrada de Aguai, s/no
13630-000 Pirassununga – SP- Brasil
maupsico@uol.com.br

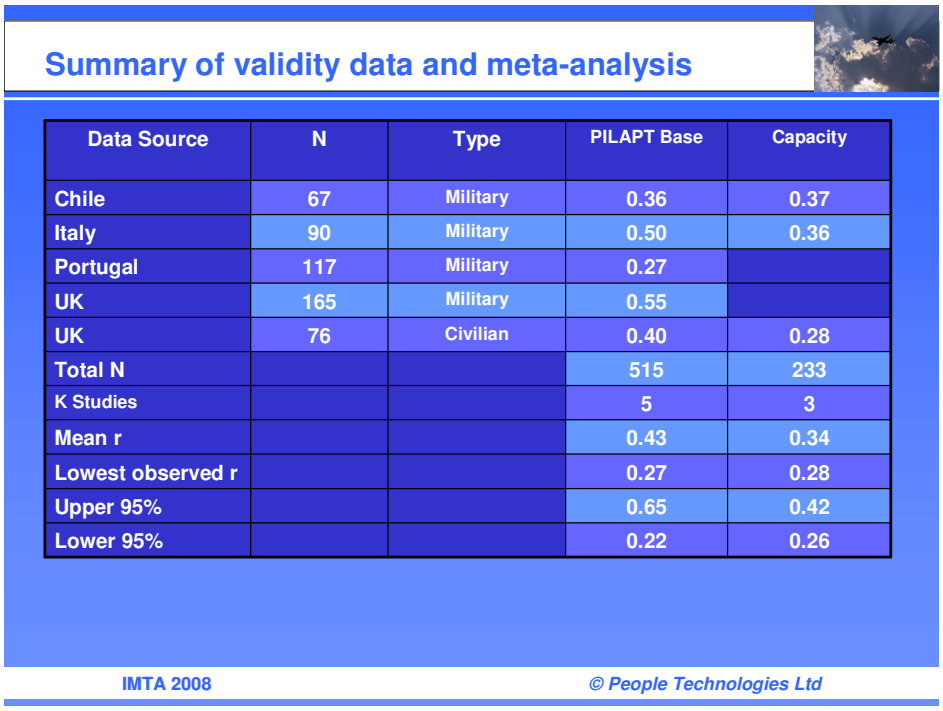
Introduction. This paper describes the results of a validation study conducted for the Brazilian Air Force using the Pilot Aptitude Tester or PILAPT. PILAPT is a fully automated battery of tests that comprises two sub batteries as follows:

- **PILAPT Base** comprises five tests that assess the basic abilities identified by Hunter and Burke (1994) as consistent predictors of success in civilian and military pilot training programmes. The tests are:
  - Deviation Indicator (DI), a measure of two-dimensional compensatory tracking
  - Trax, a measure of three-dimensional pursuit tracking
  - Hands, a measure of perceptual speed (speed and accuracy of information processing) and spatial orientation
  - Patterns, a timed measure of perceptual closure or the ability to ignore distractions in identifying the presence or absence of a visual target
  - Concentration, a measure of time sharing and divided attention

- **Capacity** is a dedicated battery of tests that together measure a candidate's capacity to operate under increasing workload. This test begins by introducing the candidate to a single tracking (primary) task. The candidate is then trained in two additional tasks, one auditory and one visual. In the dual task condition, the candidate undertakes the primary task alongside each of the secondary tasks separately. In the triple task condition, the candidate undertakes all three tasks simultaneously. This test is used in addition to PILAPT Base by some military organisations to identify fast jet potential, and is used by civilian organisations to select direct entry candidates. Validity studies have shown strong predictions of simulator performance.

Previous research has shown the validity of these tests across various settings. The question this paper addresses is whether a new study conducted for the Brazilian Air Force is consistent with previous validation research. This represents a confirmatory approach to validation where the key question is whether there is evidence that a battery can be transported to a new setting where conditions such as applicant demographics, selection processes such as degree of pre-selection, and training criteria may be different. As such, rather than simply conduct a separate and additional validity, data from prior studies is used to set a baseline for establishing whether conditions for transportability have been met or not.

Previous validation research. Figure 1 below summarises the results obtained from validations conducted in Europe (Italy, Portugal and the UK) and in South America (Chile) for both civilian and military organisations. Data were available for 515 candidates across five studies for PILAPT Base and for 233 candidates across three studies for Capacity. A full meta-analysis for this data would not be appropriate given the number of studies involved (general recommendations are for the number of studies, K, to be at least 21). However, weighted means and upper and lower 95% confidence intervals can be created using the variance in validities across studies to establish priors against which new data can be compared. Where new data exceeds the lower 95% confidence interval, then some evidence is offered for transportability across settings. Where new data does not meet this criterion, then evidence is obtained that the test battery may not be transportable in that different conditions may result in different and possibly lower validities which would need to be investigated further.



Data Source	N	Type	PILAPT Base	Capacity
Chile	67	Military	0.36	0.37
Italy	90	Military	0.50	0.36
Portugal	117	Military	0.27	
UK	165	Military	0.55	
UK	76	Civilian	0.40	0.28
Total N			515	233
K Studies			5	3
Mean r			0.43	0.34
Lowest observed r			0.27	0.28
Upper 95%			0.65	0.42
Lower 95%			0.22	0.26

Figure 1: Summary of validity evidence prior to Brazilian study

To meet the criterion for transportability just described, validities observed in the new study would need to exceed a lower value of 0.22 for PILAPT Base and 0.26 for Capacity. The mean sample weighted validities observed from previous studies were 0.43 for PILAPT Base and 0.34 for Capacity. All correlations shown in Figure 1 are uncorrected and the variance in validities can reasonably be taken to reflect differences in conditions across candidates, selection and training contexts.

The Brazilian Air Force Study. Data were obtained from 224 candidates entering initial pilot training. The PILAPT tests were administered before entering training and candidates were then tracked through all training stages. As such, the design of the study was a predictive criterion validation rather than a concurrent (postdictive) validation. Candidates are assessed across four major criteria of performance including pre-solo (initial contact with aircraft), manoeuvres and acrobatics, formation and navigation. The quality of the criteria are strong in that candidates are observed by multiple instructors across several missions per element of training, and instructor observations are recorded using structured ratings. Performance across all missions and all sorties are averaged to provide an overall measure of training performance referred to as Flight Mean. This was used as the principal criterion for evaluating the validity of the PILAPT tests in this study.

Results. Validities were 0.45 for PILAPT Base and 0.33 for Capacity against Flight Mean. These observed validities exceed the lower 95% limits expected from previous studies of 0.22 for PILAPT Base and 0.26 for Capacity. As such, these results offer evidence of transportability of the validity of the PILAPT batteries given that the values obtained in the new study are consistent with values that would be expected from the distribution of validities from prior studies.

Table 1 below provides a more detailed breakdown of validities against overall ratings for each training element. Of the ten validities obtained, only the correlation between PILAPT Base and Navigation is not significant at the 0.001 level using one-tailed significance tests (one-tailed tests were used as the direction of the relationship is expected to be positive and above zero, and to strengthen the statistical power of the analyses). The correlation between Capacity and Navigation can be expected given the increased workload on the candidate in this training element in commanding the aircraft across different geographical points on a mission. The non-significant correlation with PILAPT Base suggests that, although this PILAPT battery does assess for spatial abilities (spatial orientation and perceptual closure), there may be a gap in the coverage of spatial ability related to this training element. Accordingly, a new test has been developed of visualisation (the ability to understand and reason with three dimensional visual data), Views, which is currently under evaluation.

Table 1: More detailed data on PILAPT validities against Brazilian Air Force training criteria

Training Criterion	PILAPT Base	Capacity
Pre-solo	**0.41**	**0.26**
Manoeuvres and acrobatics	**0.39**	**0.22**
Formation	**0.36**	**0.35**

Navigation	**0.08**	**0.25**
Flight Mean (Overall Performance)	**0.45**	**0.33**

Note: All correlations significant at the 0.001 level, one-tailed, except PILAPT Base and Navigation

Despite the potential gap in the assessment offered by PILAPT Base, the value of the PILAPT batteries in predicting training success for the Brazilian Air Force is shown by comparing the likelihood of passing once candidates have been classified by PILAPT scores. For this analysis, an overall composite score was created by adding standardised scores across all PILAPT tests (all tests were given equal weight). These scores were then converted to the Standard Ten or Sten scale using the distribution of scores for the sample of candidates.

Table 2 below then compares the likelihood of candidates passing using odds ratios. These ratios are simply computed by taking the ratio of the proportion passing to the proportion failing for each score classification and for the base rate given by taking the proportion passing overall (78%) to the proportion failing overall (22%). The base rate for this sample is 3.5 to 1. That is, candidates entering the programme as selected by the current processes used by the Brazilian Air Force are, on average, 3.5 times more likely to pass than to fail the training programme.

In Table 2 and for the sake of brevity, Sten scores have been grouped into the lowest 30% of scores (Stens 1 to 4), the middle 40% of scores (Stens 5 and 6) and the top 30% of scores (Stens 7 to 10). As Table 2 shows, those in the bottom 30% of scores on the overall PILAPT composite (PILAPT Base and Capacity) enter the programme with only a 1.5 to 1 likelihood of passing which is less than half the average of 3.5 to 1. Those in the top 30% of PILAPT scores enter the programme with a 11.6 to 1 likelihood of passing, which is over three times the average likelihood for all candidates (3.5 to 1) and almost eight times more likely to pass training when compared to the lowest 30% of PILAPT scores (1.5 to 1).

Table 2: Likelihood (odds ratio) of passing training by PILAPT score band

Classification of candidates	Pass % (P) versus fail % (F)	Odds ratio (P : F)
All candidates	**78% to 22%**	**3.5 to 1**
PILAPT Stens 1 to 4 (lowest 30%)	**60% to 40%**	**1.5 to 1**
PILAPT Stens 5 & 6 (middle 40%)	**81% to 19%**	**4.2 to 1**
PILAPT Stens 7 to 10 (highest 30%)	**92% to 8%**	**11.6 to 1**

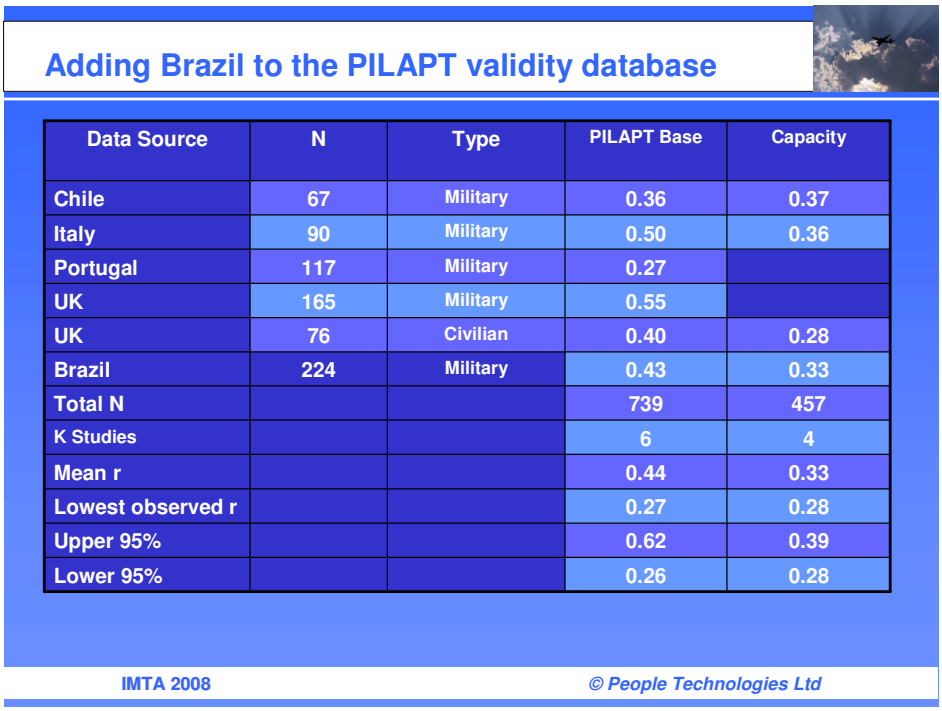
Summary. This paper has described a confirmatory approach to the validation of a pilot test battery to determine whether there is evidence for the transportability of its validity to new and different settings. Data from a study conducted on 224 Brazilian Air Force pilot candidates were compared to prior validity evidence obtained from studies conducted in Europe and South America for civilian and military organisations. The results suggest evidence for transportability of PILAPT validities. A more detailed examination of the data obtained from the Brazilian study has prompted further PILAPT research and development to assess visualisation ability, but the current tests also show high levels of prediction of success in Brazilian Air Force pilot training. Indeed, those scoring in the highest 30% of PILAPT scores were found to be three times more

likely to pass training than the average of all candidates, and almost eight times more likely to pass training when compared to candidates scoring in the lowest 30% of PILAPT scores.

Of course, the Brazilian study becomes an addition to the PILAPT validity database which can then be used to determine prior values for future validation studies. Figure 2 below provides a summary of PILAPT validities with the Brazilian data added.

References.

Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot training success: A meta-analysis of published research. *International Journal of Aviation Psychology*, **4**, 297-313.



Data Source	N	Type	PILAPT Base	Capacity
Chile	67	Military	0.36	0.37
Italy	90	Military	0.50	0.36
Portugal	117	Military	0.27	
UK	165	Military	0.55	
UK	76	Civilian	0.40	0.28
Brazil	224	Military	0.43	0.33
Total N			739	457
K Studies			6	4
Mean r			0.44	0.33
Lowest observed r			0.27	0.28
Upper 95%			0.62	0.39
Lower 95%			0.26	0.28

Figure 2: Summary of validity evidence with Brazil data added