



HAL
open science

Learning flight procedures by enacting and receiving feedback

Nadine Matton, Camille Vrignaud, Yves Rouillard, Julie Lemarié

► **To cite this version:**

Nadine Matton, Camille Vrignaud, Yves Rouillard, Julie Lemarié. Learning flight procedures by enacting and receiving feedback. *Applied Ergonomics*, 2018, 70, pp. 253-259. 10.1016/j.apergo.2018.03.009 . hal-01742404

HAL Id: hal-01742404

<https://hal-enac.archives-ouvertes.fr/hal-01742404>

Submitted on 26 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning flight procedures by enacting and receiving feedback

Matton, Nadine, ENAC & CLLE, University of Toulouse, France
7, avenue Edouard Belin, 31055 Toulouse Cedex 4, France
nadine.matton@enac.fr

Vrignaud, Camille, ENAC & CLLE, University of Toulouse, France

Yves Rouillard, ENAC, University of Toulouse, France

Lemarié, Julie, CLLE, University of Toulouse, France

Corresponding author : Nadine Matton

Abstract

Learning flight procedures is part of any pilot training. The conventional learning method consists in learning and practicing the procedure written on a sheet of paper along with printed images of the cockpit. The purpose of the present paper was to test the efficiency of a tactile interactive multimedia training tool designed to foster the self-regulated learning of flight procedures, especially through enacting relevant gestures and providing feedback. Results showed that learning with this tool did not lead to significant shorter learning times than with the conventional learning. However, on a delayed retention test in a real A320 cockpit simulator, learners of the experimental group performed the procedure more rapidly than those of the control group. Results suggested that a training tool that incites learners to perform similar gestures than those in the real environment and that provides feedback, helped learners to transform declarative into procedural knowledge.

Keywords: Procedural Learning; Multimedia Learning; Embodied cognition.

1. Introduction

Learning flight procedures is an important part of flight training from general aviation to airline pilots. Indeed, the sequence of actions that need to be performed at the various stages of the flight (e.g., before start, before taxi, engine run-up, before take-off..) has been codified for each type of aircraft in order to standardise the procedures and to reduce the cognitive load of the pilots. Thus, for the routine steps of the flight, the pilot does not need to engage in reflective processing and just has to follow validated sequences of actions that ensure safety requirements. Consequently, the pilot will have more mental resources available for dealing with all the elements that vary from one flight to another. For each aircraft, around 12 procedures are to be learnt, all by themselves, in a self-regulated way. To learn a procedure, the student pilot usually reads several times the successive steps written in the procedure manual along with printed images of the aircraft cockpit. During the initial flight training of airline pilots, student pilots learn to fly with several aircraft, three usually, thus they need to learn approximately 40 procedures during their initial practical training. To fly on a specific type of aircraft in an airline, pilots again need to learn its specific procedures. Therefore, there is an interest in improving the learning of flight procedures. It would be helpful for student pilots to benefit from conditions that facilitate learning. Indeed, if student pilots memorize more deeply each procedure, they will have more available cognitive resources during flight

and have more time left for other components of flight training (flying and non technical skills). To our knowledge, procedure or checklist learning methods have not been studied previously. Research papers on checklists focused rather on guidelines for the design of checklists (e.g., Degani & Wiener, 1993) or the benefits of the use of checklists in other domains inspired by aviation industry like in health care (e.g., Kramer & Drews, 2017). Only one research paper aimed at improving the use of checklists during flight training and highlighted the benefits of presenting feedback after checklist completion (Rantz, Dickinson, Sinclair, & Van Houten, 2009). Moreover, it has been shown that the design of the procedure itself may have an impact on its effective use (e.g., Drury, 1998; Drury, Drury Barnes & Bryant, 2017; Larock & Drury, 2003) and that new delivery technologies would not necessarily ensure the following of procedures (Drury & Johnson, 2013).

However, new digital technologies could allow the design of new learning tools for very specific needs such as the self-learning and self-training of flight procedures. In order to identify possible sources of improvement of learning procedures, we reviewed the principles that should be followed for the development of a new learning and training tool for self regulated learning of flight procedures. These principles emerge from cognitive psychology theories about skill acquisition, procedural document processing and embodied cognition. A tool, named PALP (Play And Learn Procedures) has been designed following these principles. A detailed description of this new learning tool will be given in the Material and Method section. In sum, PALP is characterized by inciting learners to perform relevant gestures (like pushing a given knob) and by providing personalised feedback at each step of the procedure. The aim of the present study was to test the efficiency of this new learning tool. The next sections will review the design guidelines that emerged from relevant cognitive psychology theories.

1.1. Learning a flight procedure: a process of skill acquisition

A cognitive model that well fits to learning flight procedures is the cognitive skill acquisition model of Anderson (1983). Indeed, the process of learning to perform a flight procedure by heart corresponds to the development of a skill. In Anderson's model (based on Fitts & Posner, 1967), skill acquisition comprises three stages. Applied to learning a flight procedure, in the first stage (called the *declarative stage*) learners encode declarative knowledge which corresponds to the list of steps of the procedure found in the flight procedure manual (e.g., "Set QNH, Display waypoints on ND1 for departure,..."). More precisely, learners need to memorize the spatial location of the various controls or buttons or displays which they will interact with, the type of interactions and the sequence of the successive interactions. At this stage, the use of knowledge is slow because it is in declarative form. The second stage is attained after several rehearsals and repetitions of the sequence of actions. At that stage (called the *associative stage*), errors are gradually more often detected and eliminated, and the memory traces of successful performance are strengthened. After further practice, the procedure is performed more automatically and rapidly. This is the third stage (called *autonomous stage*) where procedural knowledge governs the skilled performance. At this stage, the procedure is performed almost effortlessly with very few errors. Therefore, following this model, one possibility of improvement of the learning process would be to give the learners a tool that would help accelerate the transformation from declarative to procedural knowledge. This could be achieved by helping learners in memorizing the spatial location of the interactions and in strengthening the correct sequence of actions by providing feedback.

1.2. Learning a flight procedure: a procedural document processing task

As the first step of learning a flight procedure is to encode declarative knowledge of the procedure, another adapted cognitive model is the model of procedural documents processing (Guthrie, Bennett & Weber, 1991). In this model, the transformation of information from text to behaviour is composed of four steps for initial performance: (1) form a conceptual model of the performance, (2) encode the procedure from the document, (3) engage in self-testing and (4) conduct self-corrections. Previous studies have shown several ways to facilitate processing of procedural documents that focused on one of these four steps. For instance, providing a procedural schema will help form a conceptual model (Smith & Goodman, 1984; Holt et al., 1989). To enhance encoding of the procedure, processing will be optimal if the number of transformations required from the performer is minimal (Glover et al., 1987). Concerning self-testing, Guthrie et al (1991) recommended the design of “a procedural guidance system that incorporates frequent self-testing” (p.260). The fourth step, self-corrections, has to our knowledge not been tested in itself. Thus, following this model and previous empirical findings, designing an instructional tool that would help to form a visualisation of the localisations of the actions to perform, to engage in frequent self-testing and to produce self-corrections would be beneficial for student pilots.

1.3. Learning a flight procedure in an embodied cognition perspective

Finally, in order to facilitate the encoding of the sequence of actions to perform, following the embodied cognition literature (e.g., Barsalou, 2008; Wilson, 2002), enactment should have beneficial effects on learning a flight procedure. Indeed, learning a flight procedure involves not only declarative knowledge but also “knowledge concerning the motor actions that have to be applied” (Van Genuchten, van Hooijdonk, Schüler & Scheiter, 2014). Positive effects of enactment have already been shown in several contexts, even for learning non-procedural material. For instance, enactment of story through manipulation of objects was beneficial for memorizing a story by children (Glenberg, Gutierrez, Levin, Japuntich & Kaschak, 2004). Moreover, gestures helped children learn a new mathematical concept (Cook, Mitchell & Goldin-Meadow, 2008). In addition, imitation during retrieval was effective for learning object-manipulation verbs (de Nooijer, van Gog, Paas, Zwaan, 2013). A review of forty-nine empirical studies investigating the effects of learning by enacting, revealed a beneficial effect in thirty-six studies, yielding a median effect size of $d=0.51$ (Fiorella & Mayer, 2015, p.175). Therefore, one may assume that designing an instructional tool that incites learners to perform a similar sequence of hand movements to that performed in the real aircraft would strengthen the memorization of the flight procedure.

To summarize, preceding sections highlighted the importance of (i) providing feedback, (ii) helping forming a visualisation of the localisations of the actions to perform, (iii) engaging in self-testing, (iv) producing self-corrections and (v) inciting learners to perform a similar sequence of hand movements to that performed in the real aircraft. In the present paper we tested whether a procedural learning tool, PALP, designed to facilitate the processing and learning of flight procedures would lead to a more efficient learning. This tool was designed following several guidelines pointed in the literature review in order to maximise the chances of observing positive effects. Indeed, this study constitutes a preliminary study to evaluate the potential of such a tool.

1.4. Hypotheses

Given the elements of the literature review, we hypothesized that this learning tool would globally lead to a more efficient learning than the conventional learning method, i.e. with the procedure written on a sheet of paper. Thus, we compared an Experimental group who learned and practiced with PALP to a Control group who learned and practiced with the

paper procedure and printed pictures of the cockpit. More specifically, we tested following hypotheses:

H1: Learning duration should be shorter for the Experimental group than for the Control group.

H2: Retention should be better for the Experimental group than for the Control group.

H2a: number of errors should be lesser in the Experimental group than in the Control group

H2b: duration of procedure execution should be shorter for the Experimental group than for the Control group

Additionally, in order to better understand the possible difference in learning efficiency between the two groups, we tested the impact of the learning condition on several subjective variables collected at the end of the learning phase: self-confidence in the procedure execution, perceived difficulty and global intrinsic motivation. Indeed, one could expect a greater self-confidence and motivation as well as a lower perceived difficulty of the learning with the computer-based learning tool.

2. Material and Methods

2.1. Participants

Forty nine participants (age $M=23.1$ yrs, $SD=4.2$, 43 males) were recruited among pilot and technical aviation students of ENAC (Ecole Nationale de l'Aviation Civile). They were pseudo-randomly assigned to two groups ($n=26$ in the Experimental group and $n=23$ in the control group) in order to balance previous flying experience based on the holding of any piloting license (private pilot license, gliding license, microlight license,...). However, none of them had any flying experience on an Airbus A320 (the cockpit which was used for the experiment). They all volunteered and signed a written consent form before starting the experiment.

2.2. Materials

The experimental materials comprised a preliminary questionnaire and the learning material (different for each group).

The preliminary questionnaire aimed to assess flight and simulator experience. It was made of 5 questions and an A320 cockpit scheme that participants had to annotate a cockpit scheme with as many items as they could. Number of correct items was used as the previous knowledge variable. The learning materials differed for each group.

2.2.1. Material of the experimental group (PALP-based). The experimental group had to learn with a computer-based learning environment: PALP. It consisted of a self-paced tactile interactive multimedia tool, named PALP. It takes the form of the screenshot of an A320 cockpit (upper panel, flight control unit and lower panel) on which learners are to interact (by double tapping or by swiping up, down, left or right). The display was a 21.5-Inch touchscreen with a 1920x1080 resolution. Before starting the learning phase, participants were instructed how to use PALP through paper-based instructions and an interactive tutorial on the tactile screen in order to be familiarized with the five basic interaction types, (1) swiping up mostly used to represent "ON", (2) swiping down mostly used for "OFF", (3) swiping right for "check parameter" or "more", (4) swiping left for "less", (5) double tapping for "set" or "push" or "press". The general idea was to perform an interaction that was the most congruent with the actual action that has to be performed in the cockpit, following the embodied cognition principle. Participants were also informed that they could use three guiding levels labelled Level 0, Level 1 and Level 2 (see Figure 1). In Level 0 (see Figure 1, a panel), no

guidance was provided for the next action to perform. Moreover, in case of correct action, information about that action was displayed in a light blue rectangle and contained exactly the same information as in the paper version of the procedure. In case of wrong action, either nothing (when the action was not performed in the target location) or an orange circle/arrow (when a wrong action was performed in a correct area) was displayed where the learner touched the screen. Level 1 was identical to Level 0 except that the zone of the next action to perform was highlighted via a light blue circle (see Figure 1, *b* panel) but no cue was given about the action required. Level 2 had same characteristics as Level 1 except that more guidance was provided: area of interest of the next action was highlighted through a light blue disk and the action required indicated by a blue arrow as well as a description of the type of action to perform in a light blue rectangle (see Figure 1, *c* panel). Thus, for Levels 0 and 1, the verbal instruction was provided after the correct action and for Level 2, this verbal instruction was provided before the action. After an action on the screen, for the three guiding levels, a feedback was provided through a flashing green (if correct) or orange (if wrong) circle or arrow. The three guiding levels encouraged self-testing, and the feedback provided helped self-correcting, as noted in the procedural document literature.

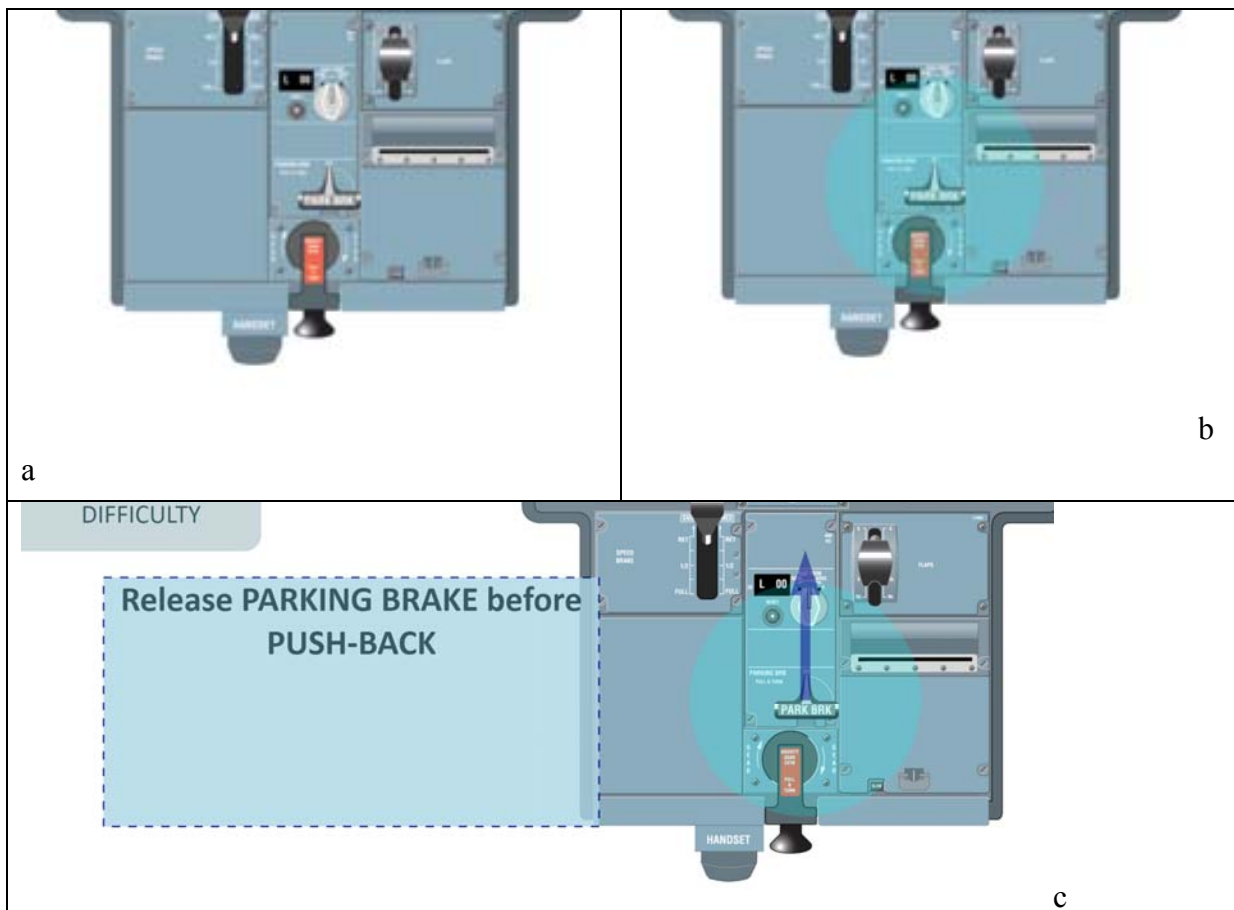


Figure 1: The three guiding levels of the computer-based learning environment before the interaction is performed: Level 0 (a) where no cue is given, Level 1 (b) where only the location of the interaction is displayed and Level 2 (c) where the action to be performed and its location are explicitly displayed.

In Levels 0 and 1, information about the action which was just completed (with exactly the same information as in the paper version of the procedure) was displayed during 3 s in a light blue rectangle after the action was correctly performed and then disappeared. In the paper-based instructions of PALP, Level 2 was explicitly recommended during exploratory phase in

order to learn the sequence of items of the procedure (it was also the default starting level). Moreover, Level 0 was recommended to be used during the practice phase. In addition, as the trainee performed the sequence of actions, an “action flow” was represented through a light blue line (see Figure 2) in order to help the trainee visually represent the whole procedure (promoting the encoding phase of the procedure).

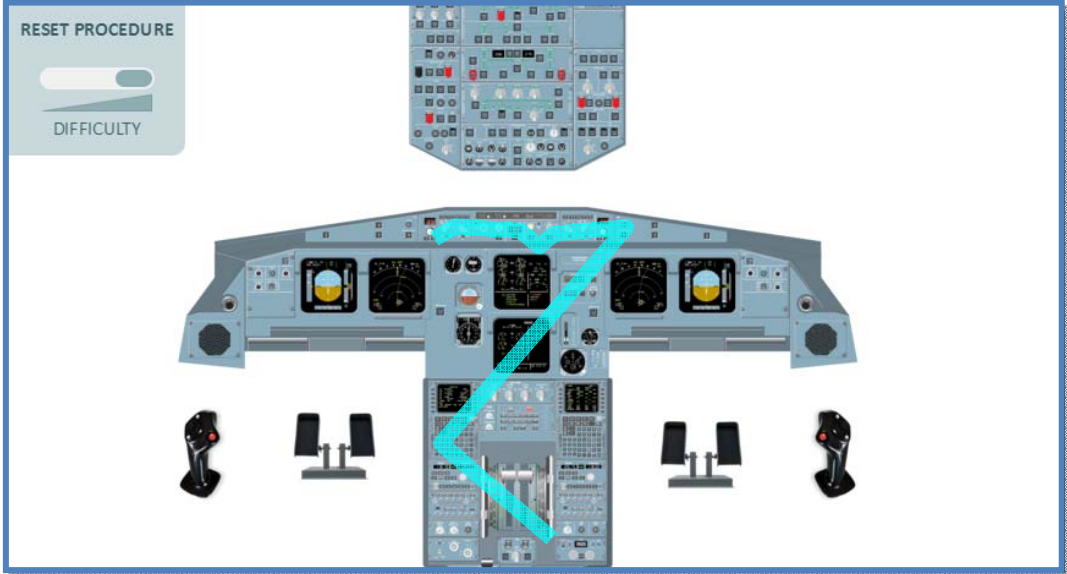


Figure 2: Blue line representing the sequence of previous actions (“action flow”).

2.2.2. *Material of the control group (paper-based).* The conventional learning environment was supposed to mimic how pilot trainees usually learn procedures. The procedure was represented in the form of a written sheet of paper listing the sequence of actions to perform (see Figure 3). They also had four size-A3 screenshots of the four parts of the cockpit (upper panel, left flight control unit, right flight control unit and lower panel). They were allowed to write on these screenshots and on the procedure sheet.

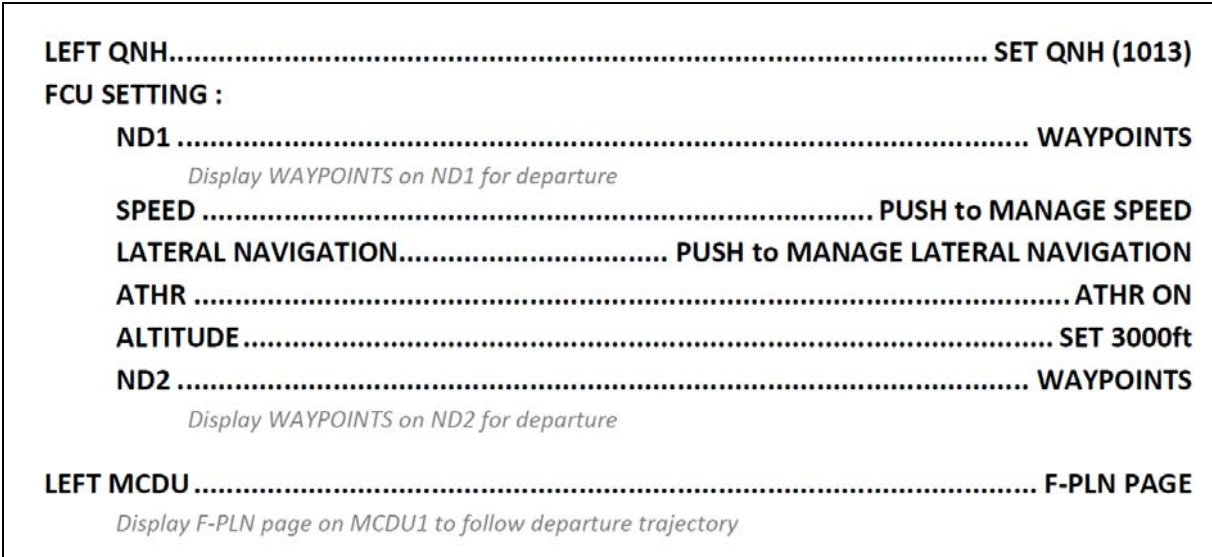


Figure 3: Eight first items of the procedure used in the conventional learning condition.

2.3. Procedure

This study used a between-subjects design. Participants were not informed of the specific objectives before the study. Only the general objective was told (“study the learning of flight procedures”) and both groups were asked to go to different rooms. Thus participants were blind to the experimental manipulation. In each group, participants had to learn the procedure composed of 22 items with no time limit. They were instructed to spend as much time as they needed until they felt confident to carry out the procedure on their own (as in Michas & Berry, 2000). Study times were analysed in the Result section and ranged from 20 to 90 mn for the Control group and from 15 to 60 mn for the Experimental group. Both groups first responded to a preliminary questionnaire in order to assess their previous knowledge and flying experience. Then they had to study a paper-based briefing about the Airbus A320 cockpit as a means to give them the basic knowledge necessary to perform the procedure (spatial location of buttons and knobs..). Participants of the control group had to record the time they started and ended the learning of the procedure. The experimental group first had to read a tutorial that explained how to use PALP and then they were invited to familiarize with the various tactile interactions. This group finally learnt the earlier mentioned procedure with PALP. Time of learning and all interactions with PALP were automatically recorded.

At the end of the learning session all participants answered to a follow-up questionnaire about their subjective assessment of the learning. An open-ended question asked them to explain the learning methods they used. They then had to rate their confidence in their capacity to execute the procedure in the simulator (1=“minimal confidence”, 10=“maximal confidence”) and their perception of the difficulty of the learning (1=“really very easy”, 7=“really very difficult”). Finally, in order to assess their motivation, they completed a 7-item questionnaire of intrinsic motivation (seven items of the 8-item questionnaire of Isen & Reeve, 2005). The seven items were as follows: “This learning stimulated my curiosity”; “This learning was interesting”; “This learning was fun”; “I want to continue learning procedures this way”; “This learning was enjoyable”; “This learning made me want to explore further the topic”; “I would be willing to come back and participate in a future experiment that used this learning device” (1=“strongly disagree,” 7=“strongly agree”).

On the next day they had to perform the procedure on a real A320 cockpit simulator at ENAC, which constituted the delayed retention test. They had to perform the procedure two times, successively, with no guidance from the experimenters concerning the steps to perform. The first run was intended to familiarize them with the A320 cockpit and the experimenters helped participants to manipulate the controls when needed (e.g., the nose wheel steering control had to be slightly pushed before moving to the “ON” position). On the contrary, during the second run, experimenters did not give any information on anything.

2.4. Analysis

Given the sample size of each group, we chose to use non parametric statistical tests in order to minimize the effect of outliers. Accordingly, for each between groups comparison we used the Mann-Whitney test. We performed statistical tests using Statistica and R software (R Core Team, 2014). To compute effect sizes we used the *compute.es* package.

3. Results

3.1. Previous knowledge and experience

Both groups had similar mean levels of previous knowledge of controls and instruments of an A320 cockpit ($M_e=6.9$, $SD_e=6.5$ vs. $M_c=6.9$, $SD_c=5.3$). Moreover, with the pseudo-random assignment to the groups based on the holding of any piloting license, a similar proportion of participants possessed a piloting license in each group (52% and 54% for Control and

Experimental groups, respectively). Therefore, previous knowledge and experience were balanced in both groups.

3.2. Time taken to study learning material until confident

One participant of the control group did not record the time at the end of the learning time, thus we had only data for 22 participants in this group. The difference of learning time between both groups was not significant ($M_e=38.4$ min, $SD_e=14.1$ vs. $M_c=45.1$ min, $SD_c=18.7$), $W=234.5$, $p=.29$, $d=0.41$. Consequently, $H1$ was not supported. However, the Experimental group spent on average 15% less time to learn the procedure than the Control group. The lack of significant difference could probably be associated to the large intra-group variability. Indeed, for the Experimental group, learning times varied from 15 min to 60 min and for the Control group they ranged from 20 min to 90 min. Interestingly, no participant of the Experimental group spent more than 60 min to learn the procedure, whereas four participants of the Control Group did (65, 78, 80 and 90 min for each of them).

3.3. Retention test (next day)

Two dependant variables were studied at each run, the number of errors and the execution duration of the procedure. Absolute numbers of errors were low. For each run, 91.8% of participants did 0 to 2 errors. This may indicate that both groups learnt seriously the procedure and/or that the procedure was easy to learn and perform. In order to compare the performance of both groups, we computed the proportion of participants who executed the procedure perfectly (see Table 1). For run#1, no significant difference between proportions of perfect execution of the procedure was observed between both groups. However, for run#2 participants of the Experimental group performed marginally better (see Table 2). Therefore, $H2a$ was not supported. For information, same results were obtained when comparing the mean number of errors ($M_c=1.7$, $SD_c=2.7$ vs $M_e=0.9$, $SD_e=0.8$ for run#1 and $M_c=1.3$, $SD_c=1.6$ vs $M_e=0.6$, $SD_e=0.8$ for run#2). Interestingly, variances of number of errors in the Experimental group were significantly smaller than those of the Control group for both runs ($F(22,25)=12.1$, $p<.001$ and $F(22,25)=4.0$, $p=.001$, respectively). More precisely, the number of errors ranged from 0 to 2 and from 0 to 3 for both runs of the Experimental group whereas they ranged from 0 to 10 and from 0 to 6 for the Control group respectively. Consequently, learning with PALP led to more homogeneous executions of the procedure in terms of number of errors than the conventional learning.

Table 1: Proportion of perfect execution of the procedure for each run. p -value of the proportion comparison test. h for effect size of the difference of the proportions.

Group	Run 1	Run 2
Exp	38.5%	57.7%
Control	26.0%	30.4%
p -value	.36	.06 [†]
h	0.27 (small)	0.56 (medium)

The Experimental group spent significantly less time to execute the procedure than the Control group for both runs, $W=162.5$, $p=.006$ and $W=163.5$, $p=.007$ (see Table 2 and Figure 4). Thus, $H2b$ was supported. On average, the Experimental group spent 31% and 21% less time than the Control group, which represented a mean difference of 38.2s and 14.5s for runs #1 and #2, respectively. Corresponding effect sizes were large ($d=0.74$ and 0.76).

Complementary analyses revealed that the duration of procedure execution was significantly positively correlated to the time spent to learn and practice the procedure, $r=.47$, $p<.001$ and $r=.36$, $p=.01$, for runs #1 and #2, respectively. Thus, on average, those participants who spent

more time to study the procedure tended to spend also more time executing the procedure. A graphical examination of corresponding correlation plots with identification of both groups highlighted that the same linear trend was observed for both groups (see Figure 5). Therefore, this trend could not be only attributed to the extreme study times obtained in the Control group. As a consequence, those participants who were more efficient during the learning phase also performed the procedure the more rapidly. The question arised whether this finding was impacted by prior piloting experience (possessing a piloting license or not). Actually, correlation between study time and execution time was non significant when computed only with participants who possessed a piloting license ($r(23)=.29, p=.15$ and $r(23)=.27, p=.19$, for runs #1 and #2, see supplementary material, Figure 7). Thus, those who had a pilot license were not necessarily those who had a shorter learning time and a shorter execution time.

Table 2: Mean duration (in seconds) of procedure execution for each run. *W* statistic and *p*-value of the Mann-Whitney test. Cohen's *d* for effect size of the difference of the means.

	Run 1		Run 2	
Group	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Exp	89.2	42.2	55.0	17.1
Control	127.5	61.1	69.5	21.3
<i>W</i>	162.5		163.5	
<i>p</i> -value	.006**		.007**	
<i>d</i>	0.74		0.76	

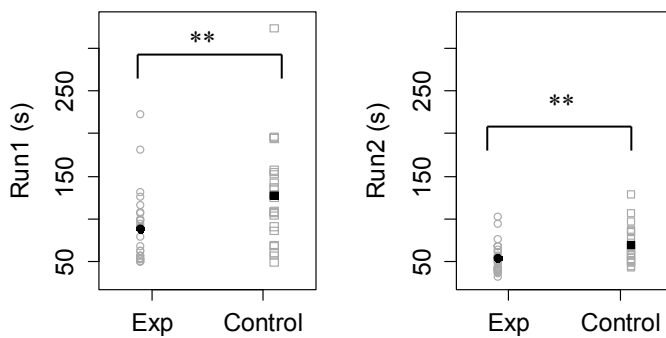


Figure 4 : Means (black filled) and raw data (not filled) of durations (in seconds) of execution for runs #1 and #2.

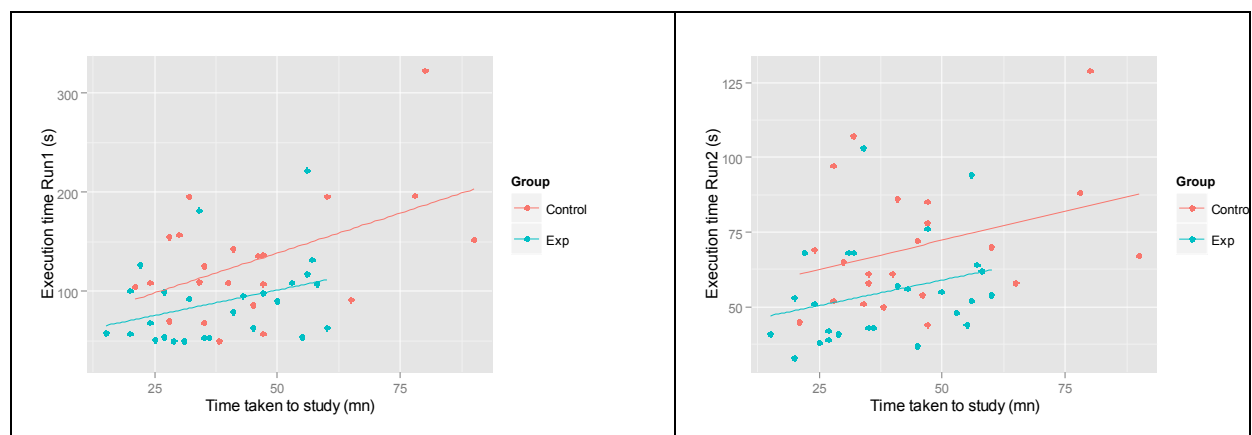


Figure 5: Correlation plots of time taken to study and procedure execution time for both groups, Control and Experimental, and for both runs (run#1 on the left and run#2 on the right).

3.4. Subjective assessments after the learning phase

No significant difference between both groups was found on any item assessed at the end of the learning phase. More precisely, no significant difference was observed for the self-confidence in the procedure execution, for the perceived difficulty of the learning or for the global intrinsic motivation (see Table 3). Thus, contrary to our expectations, learning with PALP did not induce more self-confidence or a greater self-reported motivation. However, the lack of difference in self-confidence assessment is not surprising, as both groups spent as much time as they needed until they felt confident. Finally, the learning was not considered as more difficult for the Control group than for the Experimental group.

Table 3: Mean self-reported assessments at the end of the learning phase (the score of intrinsic motivation was computed as the average of the seven scales of this questionnaire). *W* statistic and *p*-value of the Mann-Whitney test. Cohen's *d* for effect size of the difference of the means.

Group	Self-confidence		Perceived difficulty		Intrinsic motivation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Exp	7.7	1.1	2.9	0.9	5.6	0.7
Control	7.7	1.5	3.3	1.0	5.8	0.7
<i>W</i>	271.5		231.5		215.5	
<i>p</i> -value	.57		.24		.15	
<i>d</i>	0.02		0.37		0.26	

3.5. User Experience of the Experimental group

Some exploratory analyses were conducted based on the recordings of all the interactions performed by the twenty-six participants of the Experimental group in order to describe the use of the learning tool. The proportions of actions performed in the three various guiding levels were significantly different, $\chi^2(2)=33.9, p<.001$ (Friedman rank sum test). All participants started with the highest level of guidance, "Level 2", which was the default level when they started the learning phase. Afterwards, they were free to change the level when they decided to and as many times as they wanted. All participants changed at least one time the level of guidance ($M=9.9, sd=11.6$, for the number of guidance changes within a procedure). However, the large majority (65.8% on average) of actions was performed with the lowest level of guidance "Level 0" (see Figure 6). Wilcoxon tests, applied with the Bonferroni correction for multiple comparisons, confirmed that the proportion of actions performed in "Level 0" was significantly greater than actions performed in "Level 1" and "Level 2", $W=635, p<.001$ and $W=633, p<.001$, respectively. In other words, PALP users spend the majority of their learning time with the minimum level of guidance (in "Level 0" participants received no cue of the action to perform, received positive feedback when their action was correct and no cue was given when their action was wrong). Thus, PALP was essentially used for practicing the procedure and checking that the sequence of actions to perform was memorized.

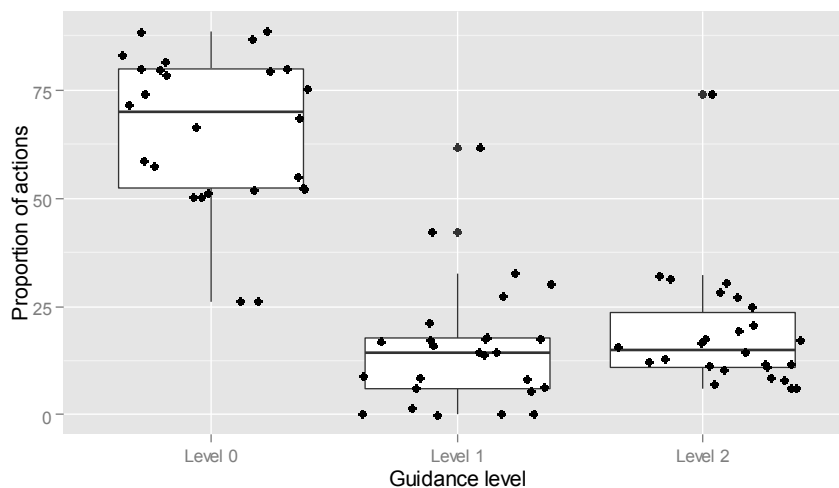


Figure 6: Box plot and raw data (black points) of proportion of actions performed in each guidance level for each participant (“Level 0” = minimum guidance; “Level 2” = maximum guidance).

3.6. User Experience of the Control group

For the Control group, we collected some information about how they proceeded. Five participants out of 23 annotated the poster: all of them numbered the actions to perform and most of them wrote a cue of the action to perform. In the follow-up questionnaire, four of them reported that they practiced by performing the actions on the poster.

4. Discussion

Traditionally, student pilots learn flight procedures by reading and memorizing the sequence of actions to perform. Following various cognitive psychology models, learning flight procedures should be fostered by (i) helping learners memorise spatial locations of the interactions, (ii) strengthening the correct sequence of actions by providing feedback, (iii) allowing to form a visualisation of the locations of the actions to perform, (iv) inciting them to engage in frequent self-testing and produce self corrections and (v) encouraging them to perform similar sequences of the hand movements that have to be performed in the real aircraft. The present study aimed at testing whether a tool respecting these design principles would lead to a more efficient learning than the conventional method.

Results showed that learning with this tool did not lead to significant shorter learning times than the conventional learning (*H1* was not supported). However, the time taken to study the material varied widely within each group (by a factor of 4 and 4.5 for the Control and Experimental group, respectively). Indeed, whereas both groups were comparable in terms of previous A320 cockpit knowledge and flying experience, within each group we had large individual differences regarding these variables. In addition, potential differences in metacognitive knowledge and strategies might also have contributed to individual differences. Our sample sizes were too small to introduce a previous experience factor in the analyses. However, for future research it would be interesting to compare times taken to learn by novices and by experimented learners.

Results on the delayed retention test in a real A320 simulator highlighted no significant mean difference between groups for the number of errors during the procedure execution (*H2a* was not supported). However, the highest numbers of errors were found in the Control group and globally both groups performed the procedure quite well. This fact might come

from the procedure used in the study, as participants could take as much time as they needed to learn the procedure (which corresponds to ecological learning conditions). Moreover, the procedure was executed 20% more rapidly by the Experimental group than by the Control group, which corresponded to a large effect size ($H2b$ was supported). This finding supports the idea of a more automated procedural knowledge with the PALP learning tool. Indeed, if the procedure is executed more rapidly, one can suppose that the Experimental group relies less on declarative knowledge than the Control group (Anderson, 1983). In critical situations one can easily imagine that the time needed to execute an emergency procedure is an important factor. Interestingly, those participants who spent the less time to study the procedure, tended to also execute the procedure more rapidly. Thus, a more efficient study time may be associated with a higher proportion of automated procedural knowledge, independently of the experimental condition. Some similar findings were found in analogous domains. For instance, a meta-analysis revealed that instructional animation outperformed static pictures for the learning of procedural-motor knowledge with a large effect size ($d=1.06$ for learning outcomes, Höffler & Leutner, 2007). Another meta-analysis highlighted that computer games outperformed conventional media for another type of procedural knowledge, namely second-language learning with also a large effect size ($d=0.96$, Mayer, 2014).

However, the question of the long term effects of the learning condition remains open in our study. Indeed, further research will be needed to test whether the difference in execution times is still observed after one week for instance. Indeed, student pilots need to remember all the flight procedures on the long run. Additionally, as the experimental tool comprised several potential sources of benefit for the learning of procedures, our study does not allow to disentangle these various effects (providing feedback, encouraging self-testing, helping to construct a global visualisation of the sequence of the locations of the actions, similarity between gestures performed during learning phase and the real situation). Indeed, as stated in the Introduction, this was an initial study to evaluate potential benefits of such a tool. As we could not predict the impact of each component, we first decided to test the tool with the full potential of benefits. Further research would be needed to estimate the relative impact of each component on the learning efficiency.

No differences were found on the subjective assessments of both groups. Thus, the difference in execution time could not be associated to a greater motivation in learning nor to a greater self-confidence nor to a lesser perceived difficulty. Moreover, study time was not longer for the computer-based learning. Thus, even behavioural cues did not indicate that students were more likely to spend more time learning flight procedures with the computer-based tool. Similar findings have been highlighted in a meta-analysis about serious games which stated that they were not more motivating than conventional instruction methods, despite positive effects on learning (Wouters, van Nimwegen, van Oostendorp & van der Spek, 2013). However, maybe differences in perseverance could be found on the long run, if students would have more than ten procedures to learn during a given period of time.

Conclusions

In conclusion, learning flight procedures with a learning tool that provides feedback, encourages self-testing, helps to construct a global visualisation of the sequence of the locations of the actions, and incites to perform gestures that are similar to those in the real situation, has proven some benefits, compared to the conventional learning. The main benefit has been found in the execution time of the procedure in a real cockpit which has been reduced by at least 20%. This means that for each procedure, the student pilot could

potentially use his cognitive resources more efficiently to other tasks needed during the flight. Further empirical findings should be collected to assess the long term effects.

Acknowledgments. We thank Matthieu Pujos, airline pilot, for the design of PALP, the learning and training tool and for the design of the A320 cockpit briefing, the PALP tutorial and the mock procedure used in this study.

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. doi:10.1146/annurev.psych.59.103006.093639

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058. doi:10.1016/j.cognition.2007.04.010

Degani, A., & Wiener, E. L. (1993). Cockpit Checklists: Concepts, Design, and Use. *Human Factors*, *35*(2), 345–359. doi:10.1177/001872089303500209

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta psychologica*, *144*(1), 173-179. doi: 10.1016/j.actpsy.2013.05.013

Drury, C. G. (1998). Case study: error rates and paperwork design, *Applied Ergonomics*, *29*(3), 213-216.

Drury, C., Barnes, C. D., & Bryant, M. R. (2017). Why do we STILL not Follow Procedures? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1664–1668. doi:10.1177/1541931213601904

Drury, C. G., & Johnson, W. B. (2013). Writing Aviation Maintenance Procedures That People Can/Will Follow. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 997–1001. doi:10.1177/1541931213571223.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks Cole.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424. doi: 10.1037/0022-0663.96.3.424

Glover, J. A., Timme, V., Deyloff, D., Rogers, M., & Dinell, D. (1987). Oral Directions: Remembering What to Do When. *The Journal of Educational Research*, *81*(1), 33–40. doi:10.1080/00220671.1987.10885794

Guthrie, J. T., Bennett, S., & Weber, S. (1991). Processing procedural documents: A cognitive model for following written directions. *Educational Psychology Review*, *3*(3), 249-265.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722–738. doi:10.1016/j.learninstruc.2007.09.013

Holt, R. W., Boehm-Davis, D. A., & Schultz, A. C. (1989). Multilevel structured documentation. *Human Factors, 31*(2), 215-228.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion, 29*(4), 295-323.

Kramer, H. S., & Drews, F. A. (2017). Checking the lists: A systematic review of electronic checklist use in health care. *Journal of Biomedical Informatics, 71*, S6–S12. doi:10.1016/j.jbi.2016.09.006

Larock, B., & Drury, C. G. (2003). Repetitive inspection with checklists: design and performance. *Contemporary Ergonomics, 477-482*.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Rantz, W. G., Dickinson, A. M., Sinclair, G. A., & Van Houten, R. (2009). The effect of feedback on the accuracy of checklist completion during instrument flight training. *Journal of Applied Behavior Analysis, 42*(3), 497–509. doi:10.1901/jaba.2009.42-497

Smith, E. E., & Goodman, L. (1984). Understanding Written Instructions: The Role of an Explanatory Schema. *Cognition and Instruction, 1*(4), 359–396. doi:10.1207/s1532690xci0104_1

Van Genuchten, E., van Hooijdonk, C., Schüler, A., & Scheiter, K. (2014). The Role of Working Memory when “Learning How” with Multimedia Learning Material. *Applied Cognitive Psychology, 28*(3), 327–335. doi:10.1002/acp.2998

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review, 9*(4), 625-636.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. doi:10.1037/a0031311

Supplementary material

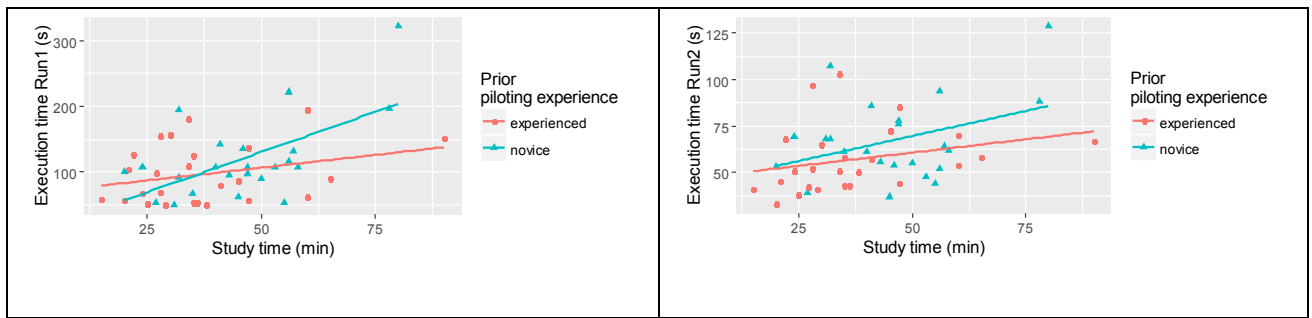


Figure 7: Correlation plots of time taken to study and procedure execution time for those participants who had a pilot license (experienced) or not (novice), and for both runs (run#1 on the left and run#2 on the right). For run#1, $r(21)=.59$, $p=.003$ for novices and $r(23)=.29$, $p=.15$ for experienced. For run#2, $r(21)=.37$, $p=.07$ for novices and $r(23)=.27$, $p=.19$ for experienced.