Training Effectiveness of Whole Body Flight Simulator Motion: A Comprehensive Meta-Analysis

Joost C. F. de Winter,1 Dimitra Dodou,1 and Max Mulder2
1Department of BioMechanical Engineering, Delft University of Technology, Delft, The Netherlands
2Control and Operations Department, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands

We present a meta-analysis of 24 effect sizes from transfer of training experiments using whole body simulator motion as an independent variable. Three moderator variables were investigated: experiment design, task type, and subjects’ experience. Due to the large heterogeneity of the included experiments, we used a random-effects model. Correction for measurement error was applied. The results revealed an overall transfer effect in favor of motion \((d = 0.51)\). Effects were weaker in true transfer \((d = 0.10)\) than in quasi-transfer with different \((d = 0.73)\) and identical \((d = 1.19)\) motion, stronger for helicopter \((d = 0.86)\) and disturbance tasks \((d = 0.84)\) than for maneuvering fixed-wing aircraft \((d = 0.07)\), and stronger for subjects without flight experience \((d = 1.57)\) than for pilots with intermediate \((d = 0.53)\) and expert \((d = -0.01)\) experience. In conclusion, motion seems important for flight-naive individuals learning tasks with external disturbances or control of vehicles with low dynamic stability, but not for experts learning fixed-wing aircraft maneuvering tasks. Effects are attenuated in true transfer as compared to quasi-transfer.

Flight simulators are valuable tools for pilot training. Training in a simulator is inherently safe and can be cost-effective. Research has shown that the use of flight simulators combined with aircraft training produces more performance...
improvements in real aircraft than aircraft-only training (J. W. Jacobs, Prince, Hays, & Salas, 1990; Orlansky & String, 1977; Pfeiffer & Horey, 1987).

Many flight simulators incorporate whole body motion feedback, typically with a 6 dof parallel manipulator, also known as a hexapod, Stewart platform, motion platform, or motion base. In contrast to the firmly established overall training effectiveness of flight simulators, the effectiveness of the motion system remains controversial. Since the early Link Trainers developed prior to World War II, during the introduction period of hydraulic hexapods in the 1970s, and nowadays with electric simulators available, there has been discussion about the training value of motion (Bürki-Cohen, Sparko, & Bellman, 2011; Caro, 1979; Cohen, 1970; Hagin, 1976; Hopkins, 1975; Martin, 1981; McCauley, 2006; Valverde, 1973). The debate on whether or not to move is relevant to flight safety. Gebman et al. (1986) found that some 15% of Air Force accidents involved situations in which motion cues were important (e.g., engine failures) and indicated that simulators with motion might better prepare pilots for dealing with safety-critical situations.

The purpose of this study is to examine the training effectiveness of whole body simulator motion. In the remainder of this introduction we briefly explain that both the performance improvements obtained in the simulator and the subjective opinions of pilots and personnel are invalid criteria for judging the training effectiveness of motion. Instead, transfer of training experiments should be used. Next, we present a comprehensive meta-analysis of transfer of training experiments with whole body motion as an independent variable. Our analysis investigated the effects of three moderator variables: experiment design, type of task, and subjects’ level of flight experience.

It should be noted beforehand that the value of motion is application-specific and that numerous other moderator variables might influence its effectiveness, including the presence and quality of the visual display, temporal synchronization between motion and visuals, the quality of auditory cues, the vehicle dynamics model and type of aircraft (e.g., fighter vs. commercial transport airplane), degrees of freedom of the motion system, duration and type of training, measurement equipment used, and the motion drive algorithm. It is also acknowledged that a large portion of the included experiments was conducted with relatively old simulator technology, and that the simulator specifications and calibration settings could have an important impact on the observed effects. The purpose of this meta-analysis is not to treat all these factors as independent variables; this would be impractical or even impossible due to the limited number of available experiments and because the necessary information is not reported in many of the studies. We do not suggest that different motion systems, such as 6 dof platforms and 1 dof tracking simulators, are equivalents. Rather, the merit of this research is to quantitatively synthesize the available information on full body motion and training effectiveness more comprehensively than previous research, and to detect
regularities of theoretical importance in terms of the three moderator variables mentioned. For specific applications of motion, one can revert to individual experiments (see De Winter, Dodou, & Mulder, 2012 for an overview) and the previous meta-analyses and reviews (to be introduced later).

**PERFORMANCE IMPROVEMENT IN THE SIMULATOR AND OPINION OF PILOTS**

In past decades and across various domains of simulation such as rotorcraft, commercial airlines, and fighter aircraft, it has been reported that motion improves in-simulator flight performance and increases the realism of pilot behavior and performance (e.g., Bürki-Cohen, Soja, & Longridge, 1998; Douvillier, Turner, McLean, & Heinle, 1960; Gundry, 1976; Pool, Mulder, Van Paassen, & Van der Vaart, 2008; Rolfe, Hammerton-Fraser, Poulter, & Smith, 1970; Showalter & Parris, 1980; Young, 1966). However, this does not imply improved learning, as humans are well able to integrate the available information to maximize their performance. Martin (1985), for example, showed that the use of a tactual seat pan (a system providing motion cues with tactile pressure to the buttock and upper thigh areas) elicited “both performance and control behavior indistinguishable from that observed in a whole body motion environment” (p. 1189). In other words, direct concurrent motion stimuli aid the pilot in the simulator by providing additional information about the simulated aircraft he or she is controlling, but the way these stimuli are perceived and processed by the pilot does not necessarily correspond to real flight (Gundry, 1976). In fact, research has shown that augmented feedback that enhances performance during training can interfere with performance in a transfer condition, because the learner has become reliant on the supplementary information (Schmidt & Wulf, 1997).

Simulator sickness is a common issue in virtual environments. Following the sensory conflict theory, a fixed-base simulator causes simulator sickness because the sensory inputs of self-motion (i.e., lack of motion in combination with visual information) differ from the inputs known from previous experience (i.e., simultaneous motion and visual cueing). Consequently, whole body motion should alleviate simulator sickness (see Jaeger & Mourant, 2001, for confirmation in a dynamic walking simulator). For flight simulation, however, the available empirical evidence is inconclusive (Sharkey & McCauley, 1992). Although tightly coupled motion might reduce simulator sickness, nonsynchronized motion or poorly tuned motion can increase simulator sickness.

Due to their kinematic and dynamic constraints, motion systems are unable to replicate all the acceleration forces produced by real aircraft (for an overview about the motion that can be provided with a Stewart platform, see Cyrus, 1977). However, by using techniques such as acceleration onset cueing as well as angular
displacement of the platform to trade the gravity vector for a perception of acceleration (a procedure known as tilt coordination), the human vestibular, tactile, and visual systems combined can provide a convincing perceptual illusion of motion (e.g., Brown, Cardullo, & Sinacori, 1989). Not surprisingly, simulator motion usually improves immersion and acceptability ratings, and instructors and pilots tend to respond skeptically when simulator motion is turned off (Boldovici, 1992; Bürki-Cohen et al., 1998; Reid & Nahon, 1988). Perception of motion is highly flexible and sensitive to task instructions, prior expectations, and reporting bias. Jacobs (1976) found evidence for the flexibility of human perception in testing a condition that randomly reversed the direction of banking motion 50% of the time when the simulated airplane was close to zero bank angle. The purpose of the experiment was concealed to the participants. Not one of the participants noticed anything unusual, even when specifically questioned about this after completing the simulator training. Similarly, Comstock (1984) investigated the effect of direction of motion in a pitching task and found that 8 of the 10 participants could not tell whether the motion was correct or reversed on a particular trial. The marked flexibility of human motion perception is not detrimental to training effectiveness per se. On the contrary, it offers possibilities for improving the perceptual fidelity of a simulator in the absence of physically realistic motion. On the same grounds, however, subjective opinion is not justified as a criterion for learning.

Summarizing, motion facilitates in-simulator performance and has various effects on pilot opinion, perception, and well-being. To assess the possible merits of motion for learning, measures that describe improvements in pilot performance in a transfer condition should be used (Caro, 1976; Gundry, 1977).

**TRANSFER OF TRAINING EXPERIMENTS**

The transfer of training paradigm is probably the most valid means of investigating the training effectiveness of motion (Advisory Group for Aerospace Research and Development, 1980). Two types of transfer of training motion experiments can be distinguished, true and quasi-transfer (Table 1). In a true transfer experiment, a group of participants is exposed to simulator training with motion. A second group is exposed to the same training with the motion system deactivated. After training, the performance of both groups is evaluated in a real aircraft. A positive training effect of motion is confirmed when the motion-trained group performs better in the aircraft than the no-motion-trained group. Quasi-transfer of training (also called pseudo-transfer, in-simulator transfer, or simulator-to-simulator transfer) follows the same procedure as true transfer, with the difference that the transfer session is conducted in the simulator acting as a stand-in for the real aircraft. A quasi-transfer design avoids the cost, hazard, and scheduling hindrances (e.g., interruptions due to bad weather) of true transfer and offers the
TABLE 1
Experiment Design of True Transfer and the Two Types of Quasi–Transfer

<table>
<thead>
<tr>
<th>Training</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>True transfer (T)</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>Simulator with motion off</td>
</tr>
<tr>
<td>Group 2</td>
<td>Simulator with motion on</td>
</tr>
<tr>
<td>Quasi-transfer type</td>
<td></td>
</tr>
<tr>
<td>different (QD)</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>Simulator with motion off</td>
</tr>
<tr>
<td>Group 2</td>
<td>Simulator in low-fidelity</td>
</tr>
<tr>
<td>motion configuration</td>
<td></td>
</tr>
<tr>
<td>Quasi-transfer type</td>
<td></td>
</tr>
<tr>
<td>identical (QI)</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>Simulator with motion off</td>
</tr>
<tr>
<td>Group 2</td>
<td>Simulator with motion on</td>
</tr>
</tbody>
</table>

possibility of testing dangerous disturbances such as engine failures (Caro, 1976; Taylor, Lintern, & Koonce, 1993).

We further distinguish between two types of quasi-transfer. In the first type, the motion group receives training in a degraded or low-fidelity motion condition and both groups are transferred to a nominal motion condition. In the second type, motion is identical in training and transfer; this means that transfer to a different motion condition takes place for the no motion group only (Table 1). The first type of quasi-transfer might be more valid for estimating the value of motion, because it assesses the generalizability of the skills to a different condition.

PREVIOUS META-ANALYSES ON TRANSFER OF TRAINING MOTION EXPERIMENTS

There have been three earlier meta-analyses of transfer of training experiments investigating simulator motion. A meta-analysis of 45 simulator-to-aircraft transfer studies by Pfeiffer and Horey (1987) showed that combined visual and motion simulation transferred better than visual training only (mean transfer ratios of 38% and 25%, respectively), and that motion simulation transferred better than no motion simulation (mean transfer ratios of 38% and 19%, respectively). However, this meta-analysis did not address causal relationships; it was based on inference instead of direct experimental manipulation of motion. Simulators with motion generally are of higher fidelity and are accompanied by more extensive whole-task training curricula (Jacobs et al., 1990). To eliminate these and other spurious effects, a meta-analysis should include only studies with motion as an independent variable, as in the meta-analyses of Jacobs et al. (1990) and Vaden and Hall (2005).

The meta-analysis by Jacobs et al. (1990) included five true transfer experiments for fixed-wing aircraft and found that the effect of motion on transfer
of training was not significantly different from zero ($d = -0.10$). Vaden and Hall (2005) included four true transfer and three quasi-transfer experiments of fixed-wing aircraft and found a slight, also statistically insignificant effect, in favor of motion ($d = 0.16$ and $0.17$ before and after correcting for measurement error, respectively). Unfortunately, these two meta-analyses were not comprehensive: Some studies were not retrieved or were excluded on the grounds of insufficient statistics even though calculating an effect size was possible. Furthermore, new transfer experiments have been conducted since 2005.

**EXPECTED EFFECTS OF WHOLE BODY SIMULATOR MOTION**

We investigated three hypotheses using the following moderator variables: (a) experiment design, considering that the training effectiveness of a flight simulator depends on the similarity between the training and transfer phase; (b) type of task, considering that the usefulness of motion depends on what behavior is trained in the simulator; and (c) experience level, considering that the learning curve model predicts that prior experience at a task correlates negatively with the rate of performance improvement (e.g., Crossman, 1959). The three hypotheses are further introduced next.

Effect of Motion as a Function of Experiment Design

It is evident that transfer of training depends on the similarity between the training and transfer phase. Complete physical fidelity is not needed to guarantee effective training: “Few would argue, for example, that transfer is enhanced if the color of an aircraft and its companion simulator are identical” (Lintern, 1991, p. 252). Nonetheless, transfer of training requires that at least the perceptually critical similarities are maintained across training and transfer tasks (Lintern, 1991).

Simulators cannot generate all the forces that occur in real aircraft. The larger hexapods provide about 35° of pitch, roll, and yaw, and about 2 m of linear displacement. Because of the limited operating space, simulator motion cues are deliberately attenuated and usually sustain for no more than 0.3 sec. The kinematic constraints also result in false motion; that is, motion occurring in the simulator only. Examples are rotational movement during simulated linear accelerations, lateral force in coordinated roll maneuvers, and motion of opposite direction to keep the motion platform within displacement limits. Furthermore, simulators always have acceleration limits and time delays. Time delays to control inputs range from 300 msec for the older platforms used in transfer of training experiments (Gebman et al., 1986; Heintzman, 1996), 100 msec for modern hydraulic platforms, down to 20 to 30 msec for the more recently developed hexapods (Berkouwer, Stroosma, Van Paassen, Mulder, & Mulder, 2005). Experiments
have shown that training with attenuated motion or severely delayed motion has a negative effect on transfer of training (Levison, Lancraft, & Junker, 1979; Nusseck, Teufel, Nieuwenhuizen, & Bülthoff, 2008).

We hypothesize that the effect of motion is stronger in quasi-transfer experiments than in true transfer experiments. In a quasi-transfer experiment, the individuals of the motion group are evaluated in a situation familiar to them, whereas in true transfer experiments they demonstrate their skills in a real aircraft with motion that might be critically different from the simulator experience. We hypothesize that quasi-transfer experiments with identical motion show a greater transfer effect than the quasi-transfer experiments involving transfer to different motion.

Effect of Motion as a Function of the Type of Task

A distinction can be made between maneuvering motion, which is motion as a result of the pilot’s actions, and disturbance motion, which is motion arising from forces external to the pilot control loop (Buckhout, Sherman, Goldsmith, & Vitale, 1963; Gundry, 1976). Hall (1989) distinguished between continuous random disturbance motion such as turbulence, and disturbance motion arising from sudden events such as wind shear, ground effect, or an engine out situation. Simulator studies have shown that motion has a beneficial effect on participants’ flight performance during disturbance tasks because it provides more lead information about disturbances than what can be obtained visually. Maneuver motion provides the pilot with additional information about the consequences of his or her control actions and is considered useful only when the aircraft is hard to fly; for example, when the aircraft is unstable or marginally stable, or when controlling the aircraft with high frequencies or high gain (Gundry, 1976; Hall, 1989). In this study, we examine whether these effects can be generalized to transfer of training. That is, we hypothesize that motion yields larger transfer effects for helicopter maneuvering and disturbance tasks than for maneuvering of typically easier to fly fixed-wing aircraft.

Effect of Motion as a Function of Participant Experience Level

There has been some disagreement as to whether simulator motion is more important for novices or for expert pilots (see Noble, 2002, for an overview). We expect that novices are at the beginning of their learning curve and still have to learn how to respond to motion cues. In contrast, for experts it is important to refresh skills already mastered in the aircraft. Experts have reached a near-asymptotic level of performance before entering the simulator, and their performance is more robust to training influences. Accordingly, we hypothesize that larger transfer effects due to motion occur for novices than for experts.
METHOD

We conducted a literature search to identify as many as possible published transfer of training experiments with motion as an independent variable. The results of Jorna et al. (1998) and Bürki-Cohen, Go, Chung, and Schroeder (2004) were obtained on request. We included only transfer of training experiments in which a group was trained with whole body flight simulator motion and another group received the same training with the motion system off. The motion during training and transfer could be of any degrees of freedom, the (simulated) aircraft and flying tasks of any type, and participants of any experience level. We defined whole body motion as moving the entire participant, including the head. This includes not only Stewart platforms, but also more limited devices such as the GAT-2 simulator providing pitch and roll motion in a range of approximately 25° (Koonce, 1974), and the Roll Axis Tracking Simulator with angular motion in 1 dof only (Martin, 1985). Experiments on systems unlikely to stimulate the vestibular organs, such as seats providing localized pressure or vibration cues, were not included in our analysis (cf. De Groot, De Winter, Mulder, & Wieringa, 2011, for nonvestibular transfer of training experiments in a driving simulator).

Our searches yielded 35 transfer of training experiments with motion as an independent variable, of which 24 were included in the meta-analysis. Details of the individual experiments as well as a list of excluded experiments and reasons for their exclusion are provided in De Winter et al. (2012).

The experiments were categorized according to experiment design (T = true transfer, QD = quasi-transfer different, QI = quasi-transfer identical), pilot task during transfer (D = disturbance task, H = helicopter maneuvering, F = fixed-wing aircraft maneuvering), and participants’ level of flight experience (N = none, participants without flight or flight simulator experience; I = intermediate, pilot students or graduates or licensed pilots without experience in the experimental tasks and aircraft; E = expert, licensed pilots or captains with experience in the experimental tasks and aircraft).

A mean standardized difference (Cohen’s $d$) between the performance of the motion-trained group and the no-motion-trained group in the transfer session was calculated for each experiment. Effect sizes were converted using the software provided by Wilson (2002). Only measures of flight performance (i.e., instructor ratings, error scores from a target aircraft state) were used to calculate $d$; measures of behavior (e.g., amount of stick movement, reaction times to impending events) were not considered. If an overall performance score was reported, only that score was used. When multiple scores were reported, $d$ was averaged first across sessions and then across tasks.

The procedure used in this study was a random-effects meta-analysis with correction for measurement error (Schmidt & Le, 2004), which, compared to a fixed-effects model, is more conservative, and which, considering the heterogeneity of
the studies involved, can be considered more appropriate because it makes the assumption that individual experiments are estimating different effects.

**Correction for Measurement Error**

We used two strategies to correct for measurement error. First, because the vast majority of studies provided no information on measurement error, we used artifact distributions for the reliability coefficients, assuming a mean reliability of .4 (with $SD$ between experiments = .1). Second, because it is questionable whether reliability distributions are equivalent between experiments, we also corrected the effect sizes individually for measurement error. Clearly, composite scores (i.e., total scores of subtasks or scores of tracking tasks measured for long periods of time) have higher reliability than scores of unaggregated single metrics (Rushton, Brainerd, & Pressley, 1983). We assumed a reliability of .2 for instructor ratings of individual maneuvers, and .4 for instructor ratings of total scores, total flight hours, or other aggregated indexes. In case insufficient data were reported to estimate the measurement error, the first two authors estimated the extent to which the dependent variables were aggregated based on consensus judgment. For objective data, we assumed a reliability of .5 for a 3-min tracking period. The effects of objective data were adjusted with the Spearman-Brown formula to obtain a reliability corresponding to the duration of the tasks. More information about our assumed reliabilities per experiment can be found in De Winter et al. (2012).

**RESULTS**

Table 2 shows the number of included experiments per combination of experiment design, task type, and experience level. Evidently, frequencies differ considerably among cells. With the exception of the stall tasks in Martin and Waag (1978a) and Woodruff, Smith, Fuller, and Weyer (1976), the included true transfer of training experiments focused on maneuvering motion only. Many quasi-transfer experiments, on the other hand, investigated disturbance motion.

Table 3 shows the results of the meta-analysis for the overall effect and for the three moderators. The columns show the number of effects, the total sample size, the observed uncorrected $d$ statistic, the $d$ statistic corrected for measurement error by using artifact distributions ($d_{ca}$), the $d$ statistic based on effect sizes corrected for measurement error individually ($d_{ci}$), and the corresponding 80% credibility interval. The results of the two corrected effect sizes were very similar (correlation of .98 between both corrected $d$ vectors reported in Table 3).

The effect of motion was positive when combining all 24 effect sizes ($d_{ci} = 0.51$). Consistent with our first hypothesis, the training effectiveness of motion was larger for quasi-transfer experiments (QD & QI) than for true transfer
TABLE 2
Frequencies of Included Effects Per Experiment Design, Type of Task, and Experience Level

<table>
<thead>
<tr>
<th>Experiment Design</th>
<th>Experience Level</th>
<th>Disturbance Task (D)</th>
<th>Helicopter Maneuvering (H)</th>
<th>Fixed-Wing Aircraft Maneuvering (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True transfer (T)</td>
<td>None (N)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Intermediate (I)</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Expert (E)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Quasi-transfer</td>
<td>None (N)</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>different (QD)</td>
<td>Intermediate (I)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Expert (E)</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Quasi-transfer</td>
<td>None (N)</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>identical (QI)</td>
<td>Intermediate (I)</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Expert (E)</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

experiments (T; $d_{ci} = 0.10$). The positive effect of motion in quasi-transfer was substantial ($d_{ci} = 0.96$) and the 80% credibility interval was not overlapping with zero. Consistent with the second hypothesis, the training effect was larger for disturbance (D) and helicopter maneuvering (H) tasks ($d_{ci} = 0.84$ and 0.86, respectively) than for maneuvering of fixed-wing aircraft (F; $d_{ci} = 0.07$). Finally, consistent with our third hypothesis, the effect was larger for individuals without flight experience (N; $d_{ci} = 1.57$) than for pilots with intermediate (I) and expert (E) levels of flight experience ($d_{ci} = 0.53$ and $-0.01$, respectively).

DISCUSSION

This study differs from previous meta-analyses on the same topic by including considerably more experiments and by investigating the training effectiveness of simulator motion for three moderator variables. Our results underline the importance of meta-analytic thinking. Whereas all the individual transfer of training studies based on statistical significance tests failed to reveal significant differences and led to inconsistent results, the meta-analytic approach detected clear regularities (F. L. Schmidt, 1992).

First, the results showed no evidence that simulator motion improves flight performance in real aircraft. The quasi-transfer experiments revealed a large positive effect in favor of motion. This result is clearly distinct from the earlier meta-analysis of Vaden and Hall (2005), which revealed no statistically significant effects. However, the importance of the positive effect should be tempered, because quasi-transfer relies on the assumption that a simulator with motion acts
TABLE 3
Results of the Meta-Analysis Per Experiment Design, Type of Task, and Experience Level

<table>
<thead>
<tr>
<th>Effects</th>
<th>Artifacts Distributions</th>
<th>Corrected Individually</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Participants</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------</td>
<td>--------------</td>
</tr>
<tr>
<td>All</td>
<td>24</td>
<td>614</td>
</tr>
<tr>
<td>Experiment design</td>
<td>T</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>QD</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>QI</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>QD &amp; QI</td>
<td>13</td>
</tr>
<tr>
<td>Type of task</td>
<td>D</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>D &amp; H</td>
<td>13</td>
</tr>
<tr>
<td>Experience level</td>
<td>N</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. T = true transfer; QD = quasi-transfer type different; QI = quasi-transfer type identical; D = disturbance task; H = helicopter maneuvering; F = fixed-wing aircraft maneuvering; N = no experience; I = intermediate; E = expert; $d_0$ = uncorrected $d$ statistic; $d_{ca}$ = $d$ statistic corrected for measurement error by using artifact distributions; $d_{ci}$ = $d$ statistic based on effect sizes individually corrected for measurement error; 80% CI represents the 10% and the 90% credibility values. In some cases the confidence interval could not be determined due to the low sample size.

as a valid stand-in for a real aircraft. There is some evidence that in-simulator effects generalize to real flight, but this evidence is not entirely supportive (Taylor et al., 1993).

Second, this study provides support for the framework put forward by Gundry (1976) regarding maneuvering motion and disturbance motion. At Gundry’s time, only a handful of transfer of training experiments were available, so he was able to test his framework only on in-simulator performance, and not training. In accordance with his predictions, we showed that training effectiveness was larger for helicopter maneuvering and disturbance tasks than for maneuvering tasks in fixed-wing aircraft. However, credibility intervals were wide and could not be determined for the latter case. A further caveat is in order, because many of the quasi-transfer experiments featured disturbance tasks, whereas many of the true transfer experiments involved maneuvering of fixed-wing aircraft. The reason is likely pragmatic, because such disturbances as turbulence or engine failures are hard to apply in a controlled manner in a real aircraft (see Ryan, Scott, & Browning, 1978, for an exception). This confounded relationship
between experiment design and type of task makes it difficult to untangle their independent effects on training effectiveness. Also, many true transfer experiments were conducted in the 1970s and 1980s on simulators with limited motion capabilities and with substantial time delays. It should be noted that time delays in the simulator (e.g., due to computer processes, visual system delay, motion system delay) are not detrimental per se. Instead, the simulator delays should be compared with those occurring in real aircraft.

To investigate whether the age of the study was related to the observed effects, we conducted a correlational analysis (see Figure 1 for the scatter plot). The correlation was not significant \(r = .05, p = .82\) when assigning equal weight to each experiment, and \(r = .06, p = .80\) when weighting the experiments according to their sample size), indicating that effects of motion were as likely to be positive in the past as they are now. This finding alone does not prove that technological advancement did not have a causal impact on training effectiveness, but it is at least consistent with such a supposition.

In agreement with our third hypothesis, the transfer effect was larger for individuals without flight experience than for pilots with intermediate or expert levels of experience. Intuitively this makes sense, as novices are at the beginning of their learning curve and still have potential for learning to use (potentially false) motion cues. Experts, on the other hand, can be expected to already have the skills to fly an aircraft. Our findings might be complicated by an interaction between pilot experience and the types of tasks: Experiments conducted with
novices often focused on basic manual control tasks, whereas the experiments involving experienced pilots tended to focus on training in more advanced disturbances such as aircraft landing and take-off with engine failures, which might be harder tasks to learn. Furthermore, expert pilots of modern jet aircraft are not necessarily experienced in all facets, because the aircraft of today are typically flown using automation, with the pilot acting as a supervisor rather than as a manual controller. This suggests that motion could be also important for experts, when learning manual control disturbance tasks.

There are several reasons why the reported effects might have been under- or overestimated. In some studies (Gray & Fuller, 1977; Martin & Waag, 1978a, 1978b; Pohlmann & Reed, 1978), the transfer phase included more than one session. It is well documented that the effect size of motion is strongest in the earlier transfer sessions and that the effect levels off in later sessions (DeBerg, McFarland, & Showalter, 1976; Jacobs et al., 1990; Levison et al., 1979; Martin, 1985). Our meta-analysis was based on the average effect of the individual transfer sessions, therefore underestimating the maximum obtainable effects, and overestimating the long-term retention effects (see Figure 2 for an illustration). The training procedures might also have contributed to the results: There is the option to train the participants to proficiency or to provide them with a fixed number of trials. Waag (1981) discussed the consequences of both approaches: Training to proficiency levels out individual differences, but the transfer variables are confounded by training time; a fixed number of trials, on the other hand, increases the variability of the aircraft performance and thus reduces the power of the design. Woodruff et al. (1976), Levison et al. (1979), and Pfeiffer and Scott (1985) used training to proficiency, whereas the other transfer studies followed the fixed training method. Furthermore, for experiments reporting multiple effects, we calculated the average $d$ among the items instead of a composite score effect. Because raw data were rarely available, we could not calculate composite scores from the individual items. Considering that each item measures only a portion of pilot proficiency, we underestimated the true effect.

Finally, there were some administrative and scheduling issues. In a number of transfer experiments, the no motion group received some training trials with motion (Woodruff et al., 1976), or participants flew in real aircraft between the experimental sessions (Koonce, 1974). System failures, instructors not following procedures, and bad weather could incidentally lead to “re-flyss” resulting in varying repetitions among students. For discussions on such limitations see Koonce (1974), Martin and Waag (1978a), Ryan et al. (1978), Woodruff and Smith (1974), and Woodruff et al. (1976).

In fighter aircraft, for conflicting or disorienting situations (e.g., high G-forces, prolonged turns), pilots are taught to ignore motion and rely on their instruments, because in such cases the human vestibular system has limited accuracy and can provide misleading illusions. This raises an interesting question: “Do we need
FIGURE 2  Cohen’s $d$ (uncorrected) between average root mean square tracking error of the motion group ($n = 6$) and the no motion group ($n = 6$) during the quasi-transfer experiment of Martin (1985; data taken from Figures 21 and 24). The subjects of the motion group were provided with whole body motion during Sessions 1 through 30. The subjects of the no motion group were provided with a static system during Sessions 1 through 20 and with whole body motion during Sessions 21 through 30. Each session consisted of four runs, 3 min each, with about a 1-min break between runs. The transfer sessions were conducted the day following the final training session. It can be seen that during training, there was a performance advantage of the motion group, which grew from about $d = 3$ to $d = 9$, a very large effect. This increase of effect is attributed to the reduced interpersonal variability with practice, resulting in reduced standard deviations of the group. During transfer, the advantage of motion gradually diminished. In our meta-analysis, the average $d$ among the 40 transfer runs was taken.

to simulate motion to learn to ignore the motion?” (McCauley, 2006, p. 10). There is evidence that the answer to this question is yes: Reed (1977) investigated visual-motion conflict while controlling remotely piloted vehicles from an airborne control station and found that for learning to disregard motion cues, it is important that motion cues are present during training.

In our literature search, strikingly few studies actually measured motion, for example, by means of accelerometers in the simulator cabin. With the exceptions of Bürki-Cohen et al. (2004), Martin (1985), and McDaniel, Scott, and Browning (1983), none of the flight simulator motion experiments included in our meta-analysis provided an accurate description of the motion hardware and motion drive laws. In fact, in aviation, there are no formal guidelines for simulator motion cueing settings, and the gains and filters of motion platforms are adjusted by trial and error by “motion tuning” experts and pilots. Fortunately, efforts are underway
to objectify and standardize the motion cueing algorithms in civil aviation simulators (Advani & Hosman, 2006). These efforts are expected to generate a major impetus to future scientific research.

As mentioned earlier, this meta-analysis focused on transfer of training, not on performance in the simulator as such. Motion might be valuable for in-simulator activities such as equipment design, pilot assessment, research on human performance, and licensing and certification. For example, a study of Schroeder and Chung (2000) showed that motion is valuable for handling qualities assessment: With large motion, pilot-induced oscillation (PIO) and handling qualities ratings corresponded to real flight data more closely than limited amplitude motion or no motion did. The work of Koonce (1974), on the other hand, indicated that not every motion is beneficial to pilot assessment. Koonce used 90 licensed pilots to evaluate three conditions in a flight simulator: no motion, simple sustained linear motion, and a more sophisticated washout motion. The predictive validity between pilot performance in the simulator and in the aircraft was .76 for no motion, .91 for sustained motion, and .65 for washout motion. Koonce concluded that for predictive validity it is important to consider which motion system provides the most consistent performance, rather than which system provides the “best” fidelity of motion or the best in-simulator performance. In-simulator performance can also be facilitated by alternative forms of motion cueing, such as motion seats, force cueing devices, and limited amplitude motion (e.g., De Groot et al., 2011; Horey, 1992; Lee & Bussolari, 1989; Martin, 1985).

The number of available experiments was limited and many of the combinations in Table 2 featured only one or even zero experiments. To obtain a better understanding of the effects of experiment design and the type of task, we recommend conducting new true transfer experiments. Specifically, we recommend replicating the helicopter experiment by Feddersen (1962). In this true transfer experiment, flight-naive participants were trained in the simulator (one motion group and one no motion group) and then had to hover a lightweight helicopter for six 2-min trials. They had to control pitch, roll, and heading attitude to maintain position over a point on the ground; altitude was controlled by a safety pilot. Performance was expressed as a displacement error from a central position. This experiment has high potential for demonstrating positive transfer of motion, because it combines helicopter maneuver motion and individuals with no flight experience.

It is interesting to extrapolate our findings to other domains, such as driver training simulators. Our results suggest that simulator motion is important when novices have to learn difficult disturbance tasks, such as terrain driving or correcting high-frequency disturbances when cornering on the verge of stability. Motion might be irrelevant to learning regular cornering, accelerating, and braking. Currently, there is no transfer of training motion research in the field of driving simulation. Most driver training simulators are far less expensive than
flight simulators, and therefore alternative motion cueing devices are an attractive research area.

In conclusion, the results of this meta-analysis suggest that whole body motion is important when flight-naive subjects need to learn helicopter maneuvering or disturbance tasks; motion might not be important for experts refreshing their maneuvering skills. More research is needed to investigate whether the skills obtained in the simulator transfer to real aircraft, and to gain insight into the independent effects of experience and type of task.

ACKNOWLEDGMENTS

The research of Joost C. F. de Winter and Dimitra Dodou is supported by the Dutch Technology Foundation (Stichting voor de Technische Wetenschappen), the Applied Science Division of the Netherlands Organisation for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), and the Technology Program of the Ministry of Economic Affairs.

REFERENCES


De Winter, J. C. F., Dodou, D., & Mulder, M. (2012). Supplementary material for “Training effectiveness of whole body flight simulator motion: A comprehensive meta-analysis.” Retrieved from https://docs.google.com/?tab=oo&authuser=0#folders/0ByG9qtADR1WHY2l0MWRjZTAtMzZmZS00YzM2LTk3ZjAiZDhiOTAxOTYxYzQw


Jacobs, R. S. (1976). *Simulator cockpit motion and the transfer of initial flight training* (Doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana, IL.


Note. References marked with an asterisk indicate studies included in the meta-analysis.

Manuscript first received: September 2010