

Fall 2002

## The Effect of Simulator Platform Motion on Pilot Training Transfer: A Meta-Analysis

Eric A. Vaden

*Embry-Riddle Aeronautical University - Daytona Beach*

Follow this and additional works at: <https://commons.erau.edu/db-theses>



Part of the [Aviation Commons](#)

---

### Scholarly Commons Citation

Vaden, Eric A., "The Effect of Simulator Platform Motion on Pilot Training Transfer: A Meta-Analysis" (2002). *Theses - Daytona Beach*. 203.

<https://commons.erau.edu/db-theses/203>

This thesis is brought to you for free and open access by Embry-Riddle Aeronautical University – Daytona Beach at ERAU Scholarly Commons. It has been accepted for inclusion in the Theses - Daytona Beach collection by an authorized administrator of ERAU Scholarly Commons. For more information, please contact [commons@erau.edu](mailto:commons@erau.edu).

THE EFFECT OF SIMULATOR PLATFORM MOTION ON PILOT TRAINING  
TRANSFER: A META-ANALYSIS

by

ERIC A. VADEN

B.S., University of Florida, 1991

A Thesis Submitted to the  
Department of Human Factors & Systems  
in Partial Fulfillment of the Requirements for the Degree of  
Master of Science in Human Factors & Systems

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Fall 2002

UMI Number: EP31881

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI<sup>®</sup>**

---

UMI Microform EP31881  
Copyright 2011 by ProQuest LLC  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

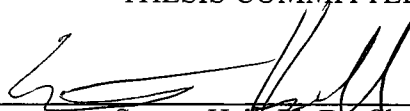
THE EFFECT OF SIMULATOR PLATFORM MOTION ON PILOT TRAINING  
TRANSFER: A META-ANALYSIS

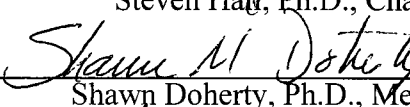
by

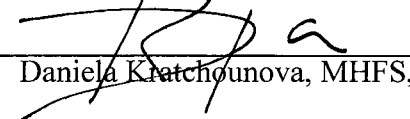
Eric A. Vaden

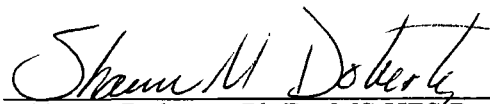
This thesis was prepared under the direction of the candidate's thesis committee chair, Steven Hall, Ph.D., Department of Human Factors & Systems, and has been approved by the members of the thesis committee. It was submitted to the Department of Human Factors & Systems and has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors & Systems.

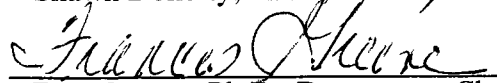
THESIS COMMITTEE:

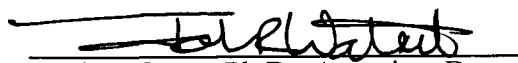
  
\_\_\_\_\_  
Steven Hall, Ph.D., Chair

  
\_\_\_\_\_  
Shawn Doherty, Ph.D., Member

  
\_\_\_\_\_  
Daniela Kratchounova, MHFS, Member

  
\_\_\_\_\_  
Shawn Doherty, Ph.D., MS HFS Program Coordinator

  
\_\_\_\_\_  
Fran Greene, Ph.D., Department Chair, Department of Human Factors & Systems.

  
\_\_\_\_\_  
John Watret, Ph.D., Associate Dean of Student Academics

## **ACKNOWLEDGEMENTS**

I would like to express special thanks to all the members of my committee for their assistance in the preparation of the thesis. Specifically, I would like to thank Dr. Steve Hall for his guidance on methodology and data analysis; Dr. Shawn Doherty for his support and enthusiasm for the topic; and Daniela Kratchounova for her optimism, continual encouragement and desire to see this project completed. Thank you all for your time and patience.

## ABSTRACT

A meta-analytic (MA) approach was used to generate an estimate of true mean effect size ( $\delta$ ) for simulator motion with regard to pilot training transfer. The analysis was based on the techniques developed by Hunter and Schmidt (1990). A  $d$  statistic was used for effect size calculations based on information available in the included sources. Eleven studies were reviewed and considered for analysis, but only seven of these included the information necessary for calculating effect size and were included in the study. The result of the MA suggest a small, positive effect for motion,  $\underline{d} = .16$ . No credibility interval could be built around this estimate of population mean effect size because the resulting sampling error variance was larger than the observed variance in  $\underline{d}$  across the assessed studies. This led to a negative variance estimate for  $\delta$  and subsequently an estimated  $\underline{SD}_\delta$  of 0. These results suggest that simulator motion has a small, positive effect on pilot training transfer and contradict an earlier MA on the same subject. The small sample size (few studies) and methodological shortcomings within the included studies require that the findings be interpreted cautiously. Alternative interpretations and their implications for the aviation training community are discussed.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
INTRODUCTION	1
Arguments For Motion	2
Arguments Against Motion	5
Previous Quantitative Reviews	8
Traditional Review Techniques	9
Meta-Analytic Approaches – An Overview	12
Anticipated Domain Specific Issues	14
HYPOTHESIS	19
METHOD	20
Setting Criteria for Study Selection	20
Literature Collection	20
Study Assessments	21
Calculating Study Effect Sizes	28
Calculations for the Bare Bones MA	29
RESULTS	31
DISCUSSION	37
CONCLUSION	43

REFERENCES	45
APPENDIX	50
STUDY SUMMARIES	50



## LIST OF TABLES

Table 1. Summary of Studies - Transfer Type, Sample Size, Trainee Experience and Simulator Type.	23
Table 2. Summary of Studies - Motion DF, FOV, Training Type and Maneuvers Assessed.	25
Table 3. Summary of Studies - Dependent Measure Type, Data Collection Technique, Analysis Type, Data Available for Calculating Effect Size.	27
Table 4. Study Sample Sizes and Effect Size Estimates.	31
Table 5. Final Values for Bare Bones Analysis.	32
Table 6. Final Values for Bare Bones Analysis Including Ryan et al. (1978).	33
Table 7. Study Values Adjusted for Reliability of the Dependent Measure	36
Table 8. Final Values for MA Corrected for Attenuation	36

## LIST OF FIGURES

Figure 1. Hypothetical Learning Curves and Pilot Performance Criteria for Motion and No-Motion in Simulator Training – Scenario 1.	42
Figure 2. Hypothetical Learning Curves and Pilot Performance Criteria for Motion and No-Motion in Simulator Training – Scenario 2.	43

## INTRODUCTION

Flight simulation has come a long way since the first Link Trainers, the famous “blue box”, the Dehmel Duplicator and the Link Translator. In the late 50’s and early 60’s, several companies incorporated motion platforms with type-specific cockpit simulators. Not only could pilots-in-training sit in and use the same cockpit layout they would experience during real flight but they could feel the simulated motion of the aircraft as well. This integration of motion has now taken the form of enormous hydraulic lift systems that afford simulated motion in all directions. Simulators built on this technology have become the status quo in high-fidelity flight simulation.

However, in the mid 1960’s, a debate began that continues to this day. That debate concerns the impact of motion in flight simulation training on training transfer. In other words, there has been a quarter century long argument over whether or not simulator motion makes any difference in the training of pilots. Hopkins (1975) was one of the earliest to argue that there was no experimental evidence in support of simulator motion when it came to enhancing pilot training transfer. He raised one of the more critical concerns in this regard, that of cost. He suggested that motion simulators that cost several times as much as the true aircraft being simulated had little or no advantage in terms of training effectiveness and might actually undermine the good use of more cost-effective simulators.

More recent authors have voiced similar concerns about the costs associated with motion platforms (see Bürki-Cohen, Soja & Longridge, 1998; Buerki-Cohen, Go, & Longbridge, 2001). In particular, Bürki-Cohen et al. (1998; Bürki-Cohen et al., 2001) cautioned against changing regulatory training requirements based on inconclusive

evidence on the effects of simulator platform motion on pilot training transfer. They suggested that regulatory changes requiring greater dependence on full-motion simulators would be especially problematic for regional airlines because of several factors including cost and availability. These authors went on to underscore a number of other critical points that contribute to the debate. Namely, they suggested the regulatory changes requiring simulator use in airline pilot training and evaluation, reduced experience levels for airline new-hires, and growing operational complexity make it necessary to review the cost effectiveness of certain simulator design attributes such as motion.

### **Arguments For Motion**

In general, those individuals supporting motion platforms have based their arguments on three main factors. First, there is a theory-based argument asserting that, in order to achieve the best training possible, and thus the greatest positive skill transfer, the training environment should be of the highest fidelity possible (Strachan, 1997; Szczepanski & Leland, 2000). Szczepanski and Leland (2000) reviewed a variety of sources to determine the necessity of motion systems for flight training in both rotary-wing and fixed-wing aircraft. They concluded that motion is necessary, particularly when the real-world task includes motion stimuli that must be interpreted accurately in order for the pilot to make proper control inputs. Specifically, they suggested that simulator motion is critical in training high G tolerance and spatial disorientation avoidance. In these tasks, they believe that visual stimulation alone from a simulator is inadequate. They argue that without an appropriate motion platform, a significant amount of information is absent from the training environment and thus training transfer may be adversely impacted. The

foundation of this argument is the century-old theory of identical elements originally posited by Thorndike and Woodworth (1901). In short, this theory suggests that the best transfer of skill from training to the operational environment will occur when the critical elements on which performance depends in the operational setting are identical in the two settings. In this case, that means that if pilot performance in the aircraft depends on motion cues and those cues can be duplicated in the training setting (the simulator), then greater transfer should occur when compared to a training environment without those cues (no motion). A host of researchers have subsequently supported and extended the basic theory of identical elements (Osgood, 1949; Holding, 1976; Anderson, 1983).

The second line of support for simulator motion comes from measures of pilot performance and control behavior during training in the simulator. Lee and Bussolari (1989) compared trainee performance under conditions of full simulator motion and special effects (small disturbance vibrations) motion only. They found that full motion cues aided student pilots in developing control strategies appropriate for the operational environment for transport aircraft while those students without full motion developed less adequate strategies. However, they did not assess transfer in their study and admit that overall performance differed little between the full motion and special effects only groups. Van der Pal (1999) found similar results when comparing full motion and no motion conditions in a quasi-transfer study (i.e., the transfer task was completed in the simulator). This author suggested that a lack motion cueing in the simulator led trainees to develop control strategies that were less successful than those developed under the motion condition when transfer was tested in a simulator under full motion conditions. This finding was specific to corrective inputs for pitch control. However, the difference

in control strategy did not affect overall performance during the transfer test.

Finally, instructor and student pilot subjective ratings of simulator training acceptance and expectations about motion effectiveness have been used to support a need for motion platforms. This support for the use of motion platforms is largely anecdotal and is generally supplied by sources considered to be subject matter experts (SMEs). Bürki-Cohen et al. (2001) reported that discussions from a series of FAA-industry symposia set up to discuss costly aspects of airplane simulation show that SMEs from industry, academia and the FAA generally believe that an absence of motion cueing in simulator platforms is detrimental to pilot control performance. The authors reported that this was particularly true for maneuvers entailing sudden motion-onset cueing with limited visual reference. Research results have also supported this line of thinking. Hall (1978) found that pilots preferred the motion to no-motion conditions when the task was to control an unstable vehicle (the maneuver studied was a Dutch roll). Ryan, Scott and Browning (1978) reported that discussions with instructors and trainees following P-3 training under motion and no-motion conditions indicated a strong preference for the use of motion cueing. They suggested, as a major conclusion in their report, that motion greatly increased pilot acceptance of the training device. Woodruff et al. (1976) reported a somewhat indirect notion of preference for motion cueing. In their study, motion cues were added to the no-motion condition when practicing a stall during T-37 training because instructor pilots believed that training without motion cueing would be ineffective. The authors admit this may have influenced the results of their motion versus no-motion comparison.

Not all preference data support the above findings. Lee and Bussolari (1989)

reported that there were no differences in instructor and trainee ratings of acceptance for full motion versus special effects only motion when the trainees were not aware of the specific motion conditions under which they trained. In an interesting twist, Jacobs and Roscoe (1975) included a randomly reversed banking motion condition in their study of simulator motion effects. In this scenario, when the trainee entered a turn, the simulator banking motion was randomized so that it may or may not have matched the turn the trainee executed. The researchers reported that not one of the trainees under the random banking motion condition commented on any odd sensations of motion and, even when asked directly, no trainee recalled experiencing motion that seemed out of the ordinary.

### **Arguments Against Motion**

Overall, empirical evidence in support of motion is lacking. Bürki-Cohen et al. (2001), in reviewing the discussions of the FAA symposia mentioned previously, indicated that, while the SMEs generally believed motion cueing to be critical, they admitted there was no scientific evidence to support such a belief. Koonce (1979) conducted a study with 90 multi-engine instrument-rated pilots participating in no motion, linear/analog motion, and full motion conditions to determine the impact of motion on the predictive validity of flight simulators for training transfer. While the no motion condition resulted in greater error in the simulator, as measured by root mean square deviation or error (RMSD or RMSE) from criteria specified in the pilot test standards (PTS), no differences were found in performance during transfer trials in the aircraft.

Jacobs and Roscoe (1975) assessed motion and no-motion conditions during

undergraduate pilot training in Singer-Link GAT-2 simulator. Using a blocked training design (i.e., all trainees received an equal amount of training) on 11 flight maneuvers, the researchers found slightly, but not statistically significant, greater transfer for a normal washout motion group versus a no-motion group. While they also reported that performance in the simulator depended on the motion condition (typically an advantage was seen for the motion group), they concluded that simulator performance and subsequent transfer performance did not show a direct relationship.

Woodruff et al. (1976) conducted a transfer of training study using motion and no-motion conditions involving the Advanced Simulator for Undergraduate Pilot Training (ASUPT) for T-37 trainees. As in the Jacobs and Roscoe study described above, no significant or practical differences were found between the motion and no-motion groups during transfer trials in the aircraft. Three more studies involving T-37 trainees (Martin & Waag, 1978a, 1978b; Nataupsky et al., 1979) also showed little evidence of a transfer benefit when using motion versus no-motion during simulator training. Ryan et al. (1978) reported similar results in their motion versus no-motion study for P-3 pilot training.

Westra (1982), using motion and no-motion simulator conditions to train carrier landings, again found no significant benefit during transfer. This study used the Visual Technology Research Simulator (VTRS) configured as a T-2c jet aircraft in a quasi-transfer design. That is, the trial used to assess the transfer of training effect was conducted in the simulator. In fact, it was conducted in the same simulator in which training took place and the motion exactly matched the motion experienced by the motion group during training. The author concluded that this implies little likelihood of seeing a



transfer benefit for motion in the real aircraft.

More recent studies show very similar results. Van der Pal (1999) assessed aerobatic and weapon delivery maneuver training and transfer in an F-16 simulator using either motion or no-motion conditions for training. Again, this was a quasi-transfer study. The author reported no evidence that motion cueing provided a benefit during training when compared to the no-motion condition. While motion tended to improve (not significantly) some aspects of control behavior (as suggested earlier), it resulted in poorer performance on other factors (e.g., absolute altitude deviation at maneuver apex). Go, Bürki-Cohen and Soja (2000) and Bürki-Cohen et al. (2001) conducted similar quasi-transfer studies with similar outcomes. In both cases, some performance measures recorded during the transfer trials showed slight benefits for motion during training (e.g., integrated airspeed exceedance) while others showed poorer performance when motion was included during training (e.g., integrated yaw activity). The researchers in both cases concluded that no operationally significant effect for simulator platform motion was apparent.

One of the few positive findings in support of simulator motion comes from the rotary wing literature. McDaniel, Scott and Browning (1983) found a positive, significant effect of simulator motion in coupled hover departure procedures while training SH-3 helicopter pilots. These authors proceeded to argue that a lack of significant motion effects in other areas should not be taken as a sign that the motion system lacks value in other operations. Only fixed wing applications are considered in the current analysis but further assessments could be made in other domains including rotary wing aircraft, marine and ground-based vehicle simulators.

## Previous Quantitative Reviews

Two prior quantitative reviews of the simulator platform motion literature have been conducted and they resulted in drastically different outcomes. Pfeiffer and Horey (1987) evaluated 45 transfer of training studies in their review effort. For each study, they computed transfer ratios (TRs) and then compared the TRs for studies that included motion in training to those that did not. The TR is indicative of the amount of training time saved in the operational setting due to prior training. In this case, it could indicate how many training flights in the aircraft might be saved by conducting prior training in the simulator. The authors reported finding strong support for the use of motion cueing based on the fact that the mean TR for studies including motion was significantly higher than the mean TR for studies not including motion. Jacobs et al. (1990) point out several problems with this argument. First, TR is influenced by the amount of training conducted. The more training you receive, the greater the TR should be. The authors do not account for this fact. Likewise, no attempt was made to weight the contribution of any given study based on sample size. Pfeiffer and Horey claimed their methodology represented a MA approach but neither the statistic being assessed (TR) nor the lack of study weightings in determining the means follows most traditional MA techniques.

Jacobs et al. (1990) conducted a MA of their own and report markedly different results. Using only studies that include motion versus no motion conditions in between-subjects designs, the researchers used calculations of point-biserial correlation ( $r_{pb}$ ) to integrate the findings of five studies. They found a small, negative effect for motion suggesting that the use of simulator platform motion might actually be detrimental to the

transfer of pilot training. However, Jacobs et al., included the results of Ryan et al. (1978) in their analysis. Their calculations produced  $r_{pb} = -0.297$  (N=50) from the Ryan et al., results. This  $r_{pb}$  was the only negative correlation coefficient of the five used by Jacobs et al. (1990), it was more than twice as large (in the negative direction) as the largest positive  $r_{pb}$  and it was weighted by the largest sample size (nearly double the next largest). All these factors caused this particular  $r_{pb}$  to have the largest impact on the final results of Jacobs et al. (1990).

Ryan et al. (1978) did not provide sufficient information to make the calculations required in the current effort and the authors did not indicate that motion had a substantial negative impact on training transfer. A calculation of effect size ( $d$ ) based on the  $r_{pb}$  reported by Jacobs et al. (1990) is included in a secondary analysis in the results section of this paper and issues regarding the inclusion of the Ryan et al. study is discussed in more detail at that point.

This very brief introduction to a quarter-century of debate is meant only to provide a backdrop to the issue of concern in this paper. The goal here is to look across the related literature of the past 25 years or more using an acceptable quantitative approach to integrate results across studies. Some typical review techniques are described in the following section.

### **Traditional Review Techniques**

Hunter and Schmidt (1990) are two of the more vocal proponents of attempts to evaluate data across studies. They argue that without such techniques, the great cumulative value of research in the behavioral sciences (and other areas) is lost.

While a variety of literature review methods have been published in the behavioral sciences, several, described briefly below, tend to dominate the literature. These prominent methods have been precipitated by the reliance on statistical hypothesis testing in the behavioral sciences. The first common review method can best be described as the voting method (e.g., Hedges & Olkin, 1980). Essentially, one would collect all the studies related to a particular research topic, hoping to include similar IV and DV comparisons, and determine the number of three possible categories of outcomes. A count would be made of positive significant effects, negative significant effects and no significant effects. The frequencies of each possibility can then be compared. If one type of outcome occurs more frequently than either of the other two, that outcome is suggested as a more accurate estimate of the true relationship between the variables under consideration. That is, it wins the vote.

Hunter and Schmidt (1990) suggested that the greatest downfall to the vote counting method is the potential for substantial levels of Type II error. Type II error occurs when a true effect exists but research results fail to identify it. Through a number of simulation tests based on distributions that assume specific true effect sizes, these authors demonstrated that some samples will produce significant results while others do not simply because of the probabilistic nature of sampling. In fact, in one example of correlational research, the authors demonstrated that, in order to achieve significance, the observed correlation must be larger than the true correlation! The authors used a Monte Carlo simulation using a true correlation of .20, study sample sizes of 40, and standard deviations of the observed (across many studies) and null distributions of .154 and .160 respectively. Based on these data, in order to be significant at the .05 level (using a one-

tailed test), the observed correlation in a given sample must be .26 ( $1.64 \times .160$ ) or greater. As the authors note, because the distribution of observed values should fall evenly about the true correlation ( $r = .2$ ,  $SD_r = .154$ ), less than half (only 35% to be exact) would fall above .26! The vote counting method would clearly not provide the correct outcome in this case since 65% of the study outcomes would not be significant (Hunter and Schmidt, 1990). While experimental rather than correlation data will be used in the current MA, Hunter and Schmidt (1990) report that the same problems with the vote counting method hold true in experimental reviews.

Another approach to integrating findings across studies might include separating the significant studies from the related but non-significant studies and attempting to find moderator variables that explain the differences in results. As Schmidt (1996) points out, the fact that some studies will result in non-significant results is easily predictable based simply on the probabilistic nature of sampling data. There is always some error that can wash out or at least attenuate effect size. Specific sources of error will be discussed later. Schmidt (1996) went on to suggest that attempting to find potentially non-existent moderators, due to the approach used above, wastes valuable research resources.

Both of the above methods have been criticized because, quite frequently, non-significant results are not published. Hence, a publication bias exists that can lead to erroneous conclusions. That is, because studies resulting in smaller, non-significant effect sizes are not often reported, they are never included in the review process. This results in the lack of a true distribution of observed effects (Hunter & Schmidt, 1990).

## **Meta-Analytic Approaches – An Overview**

In general, MA is a technique used to integrate findings across studies. In a very simplistic sense, its goal is to use data (usually an estimate of effect size) from studies in a particular research area to generate a true estimate for the effect size of a particular correlation or experimental treatment. The value in the method is that it affords scientists the ability to view findings in a cumulative form. Results of MAs can assist in the support or modification of existing theories, the definition of new theories and in the conservation of research efforts (Hunter & Schmidt, 1990).

While several methods of MA exist, only two will be described here and only at a conceptual level. One of the earliest and most widely used techniques is the Glassian approach (see Glass, 1977). The Glassian approach is generally considered a very liberal approach to MA. The first reason for this is that, according to this approach, it is valid to use multiple estimates of effect size from a single study. Hunter and Schmidt (1990) argued that this violates the fundamental rule of statistical independence and should not be allowed. That is, any study artifact (e.g., dichotomization of a continuous IV) that might produce error in the observed effect size could affect all of the effect sizes calculated (thus they are not “independent”) for a single study. Error repeated in each of the multiple effect sizes from a single study would then become overly influential in the final estimate of true effect size. It simply causes and over-weighting for some studies as compared to those from which only a single effect size can be calculated.

Further, the Glassian approach suggests that all studies in an area should be included regardless of methodological goodness. Some authors have criticized this and suggest that only those studies judged as methodologically strong should be included (see Slavin,

1986). Hunter and Schmidt (1990) supported Glass on this point because selecting only the “best” studies allows a very subjective evaluation to enter into the analysis. Finally, the Glassian approach calls for the inclusion of data from studies using a wide variety of independent and dependent variables. This point has likely resulted in the most criticism of the approach as it further enhances the liberal results of the method. Generally, this characteristic of Glassian MA has been viewed as an apples and oranges issue which increases the difficulty of interpreting the results. That is, when multiple and varied independent and dependent variables are all thrown into the mix, the final interpretation of the data will be limited (Hunter & Schmidt, 1990).

However, Hunter and Schmidt (1990) also argue two related points. First, they suggested that the studies that should be included in the analysis are dependent on the conclusion that the researcher is trying to draw. For example, if the goal is to evaluate the effect of simulator motion vs. non motion on training effectiveness, it may be quite fine to include studies using fixed and rotary-wing simulators, land-based vehicle simulators and marine vehicle simulators. Second, Hunter and Schmidt (1990) pointed out that conducting a Glassian MA does not preclude running another analysis on logical subgroups from the broader comparison. In this case, an overall analysis could be conducted first, followed by separate analyses for fixed and rotary-wing simulators.

An alternative approach has been proposed by Hunter and Schmidt (1990). In actuality, their approach is more or less a modification of the Glassian methodology. First, they allowed for only one estimate of effect size per study to protect statistical independence of the measures. Next, instead of using estimated effect sizes at face value, Hunter and Schmidt provided calculations for the variance in observed effect sizes,

$\text{Var}(d)$ , and an estimate of variance due to sampling error,  $\text{Var}(e)$ . The difference in these values is then taken as an estimate of variance in the true effect sizes,  $\text{Var}(\delta)$ . These variances can be further corrected for a variety of study artifacts such as unreliability in the dependent variable measures. Artifacts such as instrument unreliability will be described in the context of the current effort in the next section of this report.

The purpose of these variance estimates is that, quite often, variation in results across studies are mistakenly interpreted as the result of moderator variables. Hunter and Schmidt (1990) insisted that one must first consider the contribution of sampling error and other study artifacts to the overall variation across studies before making any assumptions about moderator variables. Once these corrections have been made, a credibility interval is built around the estimate of effect size using the corrected variance estimate. The size of the credibility interval then enters into the final interpretation of the results. Hunter and Schmidt advised that, when the remaining variance is small, thus leading to a narrow credibility interval about  $\delta$ , it can likely be attributed to study artifacts for which no correction is possible (Hunter & Schmidt, 1990).

### **Anticipated Domain Specific Issues**

The following paragraphs provide a more detailed description of the artifacts and other issues that were expected to have an impact this MA. For each, a brief general description is followed by a discussion of the connection the artifact may have to the present effort.

*Source Availability Bias:* Source availability bias is caused by the fact that not all studies



in a particular area of research are available for inclusion in an MA. Hunter and Schmidt (1990) suggested that certain erroneous assumptions have resulted in claims of source availability bias being the most frequent criticism of the MA approach. In general, it has been argued that unpublished studies have smaller effect sizes and are less likely to be available to be included in meta-analyses. Hunter and Schmidt (1990) pointed out that this criticism could be true of any cross-study technique including the more traditional ones described earlier in this paper. Their review of this topic included coverage of work by Rosenthal (1984) that indicated no significant difference was found between effect sizes from published and unpublished reports when 12 meta-analyses were reviewed (Rosenthal, 1984, as cited in Hunter & Schmidt, 1990).

However, Hunter and Schmidt focused most of their efforts in the organizational psychology literature and the findings reported above may not hold for human factors research. In the current effort, this issue of source availability bias seemed to be minimized. In fact, the majority of empirical evidence gathered showed null results. Thus, there appear to have been few hurdles to publishing results that show little or no effect of simulator platform motion on pilot training transfer and publication bias should.

*Data Availability:* MA procedures require particular data types from each study to be included in the analyses. In many cases, reports do not include adequate information for inclusion. Experimental studies, the most likely source of data for this effort, must include some representation of the variance accounted for by each reported effect. This could be represented by eta-squared in most reports. However, it is often omitted in final publications (Hunter & Schmidt, 1990). In the event that variance accounted for is not

reported, some other means of determining effect size must be employed. This may require making estimations or contacting the original authors.

This issue was problematic in the current effort. Very few studies were actually available for inclusion in the MA and slightly less than half had insufficient data with which to calculate effect size. One specific case, described in the results section of this report, may have significantly altered the outcome of the MA.

*Error of Measurement in the Dependent Variable:* In general, measurement error results in greater variance in performance measures and thus reduced effect size. Ideally, in the current setting, unbiased data recording could be done by the simulators themselves and data could be collected on highly reliable performance measures. In many instances, pilot performance is evaluated via subjectively scored ratings scales. These scales tend to have poor reliability both across measurements and across raters. Initially, a correction for unreliability in the performance measures based on reported reliability information was intended in the current effort. However, a lack of reporting of measurement reliability precluded such a correction. Instead, it was decided that a “worst case” scenario calculation would be made in the place of the absent reliability information. This issue is discussed further in later sections of this report.

*Error in the Treatment Variable:* Error in the treatment variable could be the result of poor measurement or poor definition. In the current domain, this may result from difficulty in measuring and defining the true motion characteristics imparted by the motion platform.

*Variations Across Studies in Treatment Strength:* In the current domain, this may result from the use of different types of simulators and different types of motion platforms. Motion is clearly not always going to be consistently applied even if it is accurately measured. Again, the small number of studies and inconsistent reporting of simulator motion properties prevented any correction relative to this artifact.

*Range Variation in the Dependent Variable:* This issue is related to the potential homogeneity in the population from which a sample comes. Individuals who participated in the studies included in this MA varied considerably across studies. It may be inappropriate to include student pilots selected for military flight programs along side student pilots who only intend to fly recreationally or even along side experienced airline pilots. This may have certain implications for the impact of simulator motion in ab initio training vs. recurrent training. Implications for this artifact are addressed in more detail in the discussion section of this paper.

*Dichotomization of the Dependent Variable:* This becomes a concern when a continuous variable is evaluated via a scale. In the specific case of dichotomization, the scale only has two points but wider scaling techniques might also attenuate effect size and reduce statistical power (Hunter & Schmidt, 1990). Data is lost any time a continuous variable is essentially turned into a categorical variable. As indicated earlier, in the current domain, it is common to find performance measurement taking the form of rating systems. Even workload measures, another common performance measure used in

aviation related studies, are often based on subjective scales. Measures such as reaction time or root mean square error may provide the most unbiased performance measures but often are not available.

*Poor Construct Validity for the Dependent Variable:* Does the measure actually capture what we think it captures? That is the critical question here. In the case of rating scale measures of pilot performance, shortcomings in the area of validity are likely. Likewise, even the less subjective performance measures may include systematic error that reduces their validity.

*Poor Construct Validity for the Independent Variable:* This issue is the result of truly confounding variables. In the current domain, one might consider how our ability to produce motion cues has changed over time. Older motion platforms did not produce the range of motion deliverable today and there was often considerable lag in the systems. Even in modern systems there may be some question about the accuracy of the motion they produce. For example, Go et al. (2000), one source of data for the current MA, admitted that their simulator may not have provided lateral acceleration cues appropriate for the maneuvers they tested during the training. While there are some techniques that can be used to correct for this fault, they are beyond the scope of this study.

*Effect Size Bias:* Hunter and Schmidt (1990) suggested that estimates of effect size that employ Cohen's  $d$  statistic tend to slightly overestimate the population effect size. They reported that the issue is of minimal consequence with sample sizes greater than 20.

Because Hunter and Schmidt generally worked in the area of organizational psychology and most of their meta-analyses dealt with correlational studies, they generally worked with studies based on larger sample sizes. However, in the current domain, sample sizes are often smaller than 20 due to the resources required to perform the experiments. A correction can be made for effect size bias in this case and a technique for that correction is presented by Hunter and Schmidt (1990).

*Recording, Computational and Transcriptional Errors:* These errors occur during the recording and transfer of data. Hunter and Schmidt (1990) identified numerous sources of such error including errors in the original data collection, errors in data entry prior to analyses and error in reporting of the analyses. Essentially, they suggested that any time numbers are worked with there is the potential for errors to be made. This type of error is often unavoidable and uncorrectable in the MA procedure.

## **HYPOTHESIS**

Ten years have passed since the last MA (Jacobs et al., 1990) was conducted in this area and more experimental data were available to include in the current effort. The MA approach reported by Hunter and Schmidt (1990) was selected to estimate the true size of the effect that simulator motion has on training transfer. This would expand the results of the Jacobs et al. (1990) MA. Given the consistent findings of the most recent studies with those of the past, it was expected that simulator platform motion would be found to have a *minimal and possibly small, negative effect on transfer of pilot training* and the results of the Jacobs et al. (1990) study would be supported.

## **METHOD**

### **Setting Criteria for Study Selection**

Several key factors influenced the selection of studies for this analysis. First, only studies involving fixed-wing aircraft training were considered. Next, only studies including simulator training with independent samples in motion and no-motion conditions were selected. The one exception to this criterion was Lee and Bussolari (1989). The “no-motion” condition in that study included bump and buffet cues for which the maximum extension of the motion platform legs was 0.25 inches. This study was not included in the final analysis however because the publication did not include adequate data with which to calculate a study effect size. Finally, only studies that included either true transfer or quasi-transfer trials were considered.

### **Literature Collection**

Searches were conducted on a variety of publication databases. Key word searches began with the general terms “simulator” and “motion”. This search was conducted on the Aerospace and High Technology Database, the database for the National Technical Information Service (NTIS), the database for the Scientific and Technical Information Network (STINET) which is the public side of the Defense Technical Information Center (DTIC) and the PsychINFO database. These searches resulted in approximately 250 hits and each associated abstract was reviewed. A contact was also made with personnel at the Marine Corps Program Directorate of the Naval Air Warfare Center Training Systems Division (NAWCTSD) in Orlando, FL. They were able to provide a wealth of potential

sources that they had collected during their research on the motion-cueing requirements for the Advanced Amphibious Assault Vehicle (AAAV) driver simulator (Jones & Franklin, 1999). Contact was also made with Ian W. Strachan who provided useful resources as well.

Roughly 70 potential sources of study data and other relevant reports were then reviewed. Reference sections of these publications were also used to identify further potential studies for inclusion in the analysis. In the end, only 11 studies were identified that met the criteria described previously. Of these 11, only seven contained sufficient information to calculate study effect sizes.

### **Study Assessments**

Research articles were reviewed and evaluated based on a few critical characteristics. Initially, it was intended that subgroups of the included studies could be created based on these characteristics and analyzed separately. However, the paucity of empirical studies meeting the basic criteria already described eliminated any opportunity for this. The primary characteristics of interest for each study were:

- Transfer technique – True Transfer or Quasi-transfer.
- Sample Size for the Motion and No-motion groups.
- Participant Experience Level
- Simulator Type
- Degrees of Freedom (DF) for the Motion Platform
- Field of View (FOV)
- Training Type – Criterion based or Blocked

- Maneuvers Assessed
- Dependent Measure Type - Subjective or Objective
- Data Collection Technique – Electronic or Hand Scoring
- Analysis Type
- Data Available for Estimating Effect Size

Complete summary sheets for each of the 11 studies reviewed are included as Appendix A. Tables 1 through 3 present the relevant information for items listed above for each study reviewed.



Table 1. Summary of Studies - Transfer Type, Sample Size, Trainee Experience and Simulator Type.

Study Reference	Transfer Type	Sample Sizes (motion/no-motion)	Participant Experience Level	Simulator Type
1. Buckhout et al. 1963	Quasi	8/8	Low hour pilots	Grumman Multipurpose Motion Sim
2. Jacobs and Roscoe, 1975	True Transfer Piper Cherokee Arrow	9/9	Low – undergraduates	Singer-Link GAT-2
3. Woodruff et al., 1976	True Transfer T-37	4/4	Low – undergraduates	ASUPT
4. Ryan et al. 1978	True Transfer P-3	39/11	Low – undergraduates	2F87F – P-3 Orion, 4 engine turbo prop.
5. Martin and Waag, 1978a	True Transfer T-37	8/8	Low- undergraduates	ASPT
6. Martin and Waag, 1978b	True Transfer T-37	12/12	Low- undergraduates	ASPT
7. Nataupsky et al., 1979	True Transfer T-37	16/16	Low - undergraduates	ASPT
8. Westra, 1982	Quasi	16/16	Mixed but no carrier landing experience	VTRS – T-2C Jet
9. Lee and Bussolari, 1989	Neither – no training just testing	6/6 8/8	2.4 year average in Exp. 1, no hours in model in Exp. 2	Boeing 727-700
10. Van der Pal, 1999	Quasi	6/6	High – retired F-16 pilots	F16

Table 1 (continued). Summary of Studies Transfer Type, Sample Size, Trainee Experience and Simulator Type.

Study Reference	Transfer Type	Sample Sizes (motion/no-motion)	Participant Experience Level	Simulator Type
11. Go et al. 2000	Quasi	18/19 or 16/18 depending on DV assessed	High – regional airline pilots in recurrent training.	Level C, 30 passenger, twin engine, turbo prop

Table 2. Summary of Studies - Motion DF, FOV, Training Type and Maneuvers Assessed.

Study Reference	Motion DF	FOV (Horizontal x Vertical)	Training Type	Maneuvers Assessed
1. Buckhout et al. 1963	3	Not given – 4 inch CRT used as display	Blocked – 15 trials, 3 transfer trials	Tracking task, low altitude flight
2. Jacobs and Roscoe, 1975	6	Not given	Blocked – trials not given	11 overall but specifics not given
3. Woodruff et al., 1976	6	Not given	Criterion – followed training syllabus	All in program – collapsed data for Basic, Presolo, Advanced Contact, Instruments, Formation and Navigation
4. Ryan et al. 1978	6	50 x 38	Criterion	3 and 4 engine aborts, Engine failure after refusal, Instrument Tasks, Landings.
5. Martin and Waag, 1978a	6	“Full” but no measure	Blocked – 10 sorties in ASPT	Basic Work – 12 maneuvers Pattern Work – 4 maneuvers Mission Profiles – all 16
6. Martin and Waag, 1978b	6	“Full” but no measure	Blocked – 5 then 2 sorties in ASPT	Basic Aerobatics – 4 maneuvers Advanced Aerobatics – 5 maneuvers.
7. Nataupsky et al., 1979	6	300 x 150 or 48 x 36	Blocked – 4 trials	Takeoff, Steep Turn, Slow Flight, Straight-In (before glidepath), Straight-In (On Glidepath).
8. Westra, 1982	6	160 x 80	Blocked – 40 trials	Circling approach and Landing (on simulated carrier)

Table 2 (continued). Summary of Studies - Motion DF, FOV, Training Type and Maneuvers Assessed.

Study Reference	Motion DF	FOV (Horizontal x Vertical)	Training Type	Maneuvers Assessed
9. Lee and Bussolari, 1989	6	Did not report	None	3 scenarios – flameout on takeoff, air work, ILS approach and landing with windshear
10. Van der Pal, 1999	6	142 x 110	Blocked – 20 trials aerobatics  12 trials weapons	Weapons delivery
11. Go et al. 2000	6	150-40	Criterion – followed ongoing training	Engine failure on Rejected Take-Off (RTO) or Continued Take-Off (V1 cut)

Table 3. Summary of Studies - Dependent Measure Type, Data Collection Technique, Analysis Type, Data Available for Calculating Effect Size.

Study Reference	Measure Type	Data Collection Technique	Analysis Type	Data for Effect Size
1. Buckhout et al. 1963	Objective – RMSE, time on target, altitude penetration, crashes	Collected from Sim	ANOVA	Insufficient data – Overall F reported for 8 groups of various motion types
2. Jacobs and Roscoe, 1975	Subjective – time/trials to criterion	Paper/pencil IP ratings	ANCOVA	Insufficient – p-values only
3. Woodruff et al., 1976	Subjective – time to criterion	IP Ratings	Ratio of hours to criterion	Raw data provided
4. Ryan et al. 1978	Subjective Ratings	Paper/pencil IP Ratings - UBAA	Repeated Measures ANOVA	Insufficient data – repeated measures $\underline{F}$
5. Martin and Waag, 1978a	Subjective Ratings	Paper/pencil IP Ratings -12 point scale	ANOVA for each of 16 measures	16 univariate $\underline{F}$ values
6. Martin and Waag, 1978b	Subjective Measures on Score Cards	Paper/pencil IP scoring of special score cards	ANOVA and a priori t-tests for each of 40 measures	40 univariate $\underline{F}$ values and independent samples t-tests

Table 3 (continued). Summary of Studies - Dependent Measure Type, Data Collection Technique, Analysis Type, Data Available for Calculating Effect Size.

Study Reference	Measure Type	Data Collection Technique	Analysis Type	Data for Effect Size
7. Nataupsky et al., 1979	Subjective Ratings and Measures on Score Cards	Paper/pencil IP ratings on 8-point scale and scoring of special score cards	ANOVA for each measure	Univariate $F$ values
8. Westra, 1982	Objective	Collected from Sim	ANOVA for each measure	Univariate $F$ values
9. Lee and Bussolari, 1989	Subjective and Objective	Paper/pencil IP ratings and collection from Sim	ANOVA	Few numbers provided – no good data for MA because no transfer measured.
10. Van der Pal, 1999	Objective	Collected from Sim	ANOVA for each measure	Only partial univariate $F$ values reported.
11. Go et al. 2000	Objective	Collected from Sim	$t$ -tests	Only $p$ -values given for $t$ -tests.

### Calculating Study Effect Sizes

Based on the data provided in studies 2, 4, 5, 6, 7, 8 and 9 above, study effect size estimates were calculated. All estimates were based on  $t$ -scores either directly reported in the studies, calculated from raw data available or calculated from reported  $F$  values. If sample sizes were equal, the equation used for converting  $t$  to  $d$  was  $d = 2t / \sqrt{N}$  where  $N$  represent the total sample for the variable tested. If sample sizes were unequal, the

equation used for this conversion was  $\underline{d} = (1/\sqrt{pq})t/\sqrt{N}$  where  $p$  and  $q$  are the proportion of participants in the two groups. These equations are presented in Hunter and Schmidt (1990).

If sufficient information was reported on multiple performance measures, an effect size estimate was calculated for each measure in a given study. A weighted mean effect size per study was then calculated. Weights were based on the  $N$  for each measure. If all measures included an equal sample size, the mean study effect size was simply the arithmetic mean of the effect sizes calculated.

### **Calculations for the Bare Bones MA**

Seven study effect sizes were then used for the final analysis following the bare bones MA technique developed by Hunter and Schmidt (1990). Calculations included an average study effect size  $\{Ave(\underline{d})\}$ , variance in the observed study effect sizes  $\{Var(\underline{d})\}$ , estimated variance due to sampling error  $\{Var(\underline{e})\}$ , estimated variance for the true population effect size  $\{Var(\delta)\}$  and finally a standard deviation for estimated population effect size ( $\underline{SD}_\delta$ ). The construction of a 95% credibility interval about  $Ave(\delta)$  was intended but  $Var(\delta)$  was negative and thus no credibility interval could be generated. Reasons for this outcome are discussed in subsequent sections. The equations used for these calculation included:

$$Ave(\underline{d}) = \sum w_i \underline{d}_i / \sum w_i = \underline{D}$$

$$Var(\underline{d}) = \sum w_i (\underline{d}_i - \underline{D})^2 / \sum w_i = \underline{D}$$

$$\text{Var}(e) = [(N-1)/(N-3)][(4/N)(1 + \delta^2/8)]$$

$$\text{Var}(\delta) = \text{Var}(d) - \text{Var}(e)$$

$$SD_{\delta} = \sqrt{\text{Var}(\delta)}$$

$$95\% \text{Conf. Int.}(\delta) = \text{Ave}(d) \pm 1.96SD_{\delta}$$

In the calculation of  $\text{Var}(e)$ ,  $\text{Ave}(d)$  is substituted for  $\delta$  as the effect size statistic.  $\text{Ave}(d)$  becomes an estimate of the true population effect size parameter  $\delta$ . All of these equations are presented in Hunter and Schmidt (1990).

Finally,  $\text{Ave}(d)$  was corrected for small sample bias using the equation  $d^* = d/a$  where the bias multiplier  $a = 1 + .75/(N-3)$  and  $N$  is the average sample size of the studies included in the MA. These equations are reported in Hunter and Schmidt (1990).



## RESULTS

Average study effect sizes ( $d$ ) and study sample sizes ( $N$ ) are shown in Table 4 for each of the studies included in the overall MA. Positive effect size estimates represent greater training transfer for the motion condition.

Table 4. Study Sample Sizes and Effect Size Estimates.

Study Reference	$N$	$d$
(Woodruff et al., 1976)	8	0.5425
(Martin and Waag, 1978a)	8	0.2154
(Martin and Waag, 1978b)	24	0.1242
(Nataupsky, et al., 1979)	32	0.3120
(Westra, 1982)	32	0.3476
(van der Pal, 1999)	12	0.0115
(Go et al., 2000)	36	-0.1462

Based on the seven mean study effect sizes shown in Table 4, the equations presented earlier were used to make the final calculations for the MA. The results of those calculations are presented in Table 5.

Table 5. Final Values for Bare Bones Analysis.

Variables	Value
Ave( <u>d</u> )	0.16
Var( <u>d</u> )	0.0442
Var( <u>e</u> )	0.2045
Var( $\delta$ )=	-0.1603
<u>SD</u> $\delta$ =	0.0

The negative value for  $\text{Var}(\delta)$  prevented the development of a 95% credibility interval around  $\delta$ . Hunter and Schmidt (1990) suggest that some bias can exist when studies rely on small sample sizes, particularly for sample sizes under 20. They report that the bias becomes negligible for sample sizes of 50 or more. The average sample size included in this MA was approximately 22 so the bias multiplier  $\underline{a}$  was calculated and applied to Ave(d). The corrected d (d\*) was 0.158, a very slight variation from the original d of 0.16. Therefore the bias multiplier was shown to have minimal impact and was not carried through the rest of the values presented in Table 5.

At this point, it should be noted that two of the five studies included in the Jacobs et al. (1990) study were not included in the bare bones MA reported in Table 5. It was decided that only studies from which a direct calculation of effect size was possible would be included in the current MA. The Gray and Fuller (1977, as reported in Jacobs et al., 1990) study could not be obtained and the Ryan et al. (1978) study did not include sufficient information for a calculation of effect size. The exclusion of the Ryan et al. study is particularly problematic because the point-biserial correlation ( $r_{pb}$ ) calculated by

Jacobs et al. (1990) for that study was large and negative,  $r_{pb} = -0.297$ . In fact, this correlation coefficient was the largest of any of the studies included in the Jacobs et al. MA and was also based on the largest sample size,  $N = 50$ . Using the equation  $d = \sqrt{[(N-2)/N]}(1/\sqrt{pq})r/\sqrt{1-r^2}$  (from Hunter & Schmidt, 1990) where  $r$  is the  $r_{pb}$  and  $p$  and  $q$  are the proportion of subjects in each treatment group, an estimate of effect size was calculated for the Ryan et al. study based on the  $r_{pb}$  reported in Jacobs et al. (1990). When this study effect size ( $d = -0.7357$ ) was added to the original bare bones MA reported in Table 5, the results in Table 6 were obtained. The outcome is substantially different. The overall effect for motion appears slightly negative ( $d = -0.06$ ) rather than positive and a 95% credibility interval can be built around the estimate of  $\delta$  such that  $-0.269 < \delta < 0.1526$ .

Table 6. Final Values for Bare Bones Analysis Including Ryan et al. (1978).

Variables	Value
Ave( $d$ )	-0.06
Var( $d$ )	0.1842
Var( $e$ )	0.1727
Var( $\delta$ )=	0.0115
$SD_{\delta} =$	0.1075

Note that, given Ave( $d$ ) of 0.16 and Var( $d$ ) = 0.0442 ( $SD_d = 0.2102$ ) for the seven studies included in the original MA, the study  $d$  of -0.7357 is 4.26 standard deviations below Ave( $d$ ). This would be a surprising outcome given that Ryan et al. (1978) do not report any substantially negative trends in training transfer as a result of their motion

treatment. They report that, for the five maneuvers believed to be most affected by motion cueing in their study, trials to proficiency in the aircraft did not differ significantly for the motion and no-motion training groups. Given the large, negative effect used for calculations in the Jacobs et al. (1990) MA, Ryan et al. (1978) surely would have reported strong negative trends for the motion group even if they could not show a significant difference between motion and no-motion. Because they report no such negative trends and because the data available in Ryan et al (1978) preclude the direct calculation of  $\underline{d}$ , the exclusion of the study  $\underline{d}$  based on the data provided by Jacob et al. (1990) seems warranted.

In one final calculation, the original MA reported here was recalculated using an attenuation factor for unreliability in the dependent variables assessed. While reliability data was not available in the four studies using subjective, IP evaluations to judge performance, it was decided to show a “worst case” scenario calculation. Holt, Hansberger and Boehm-Davis (2002) provide a starting point for estimating unreliability for pilot ratings using a 4-point scale (similar to one used in some studies included in this MA). In the development and assessment of their rater training program, Holt et al. collected base-line data that suggested interrater correlation of about .56. For the recalculation of the original MA, it was decided that an IRR of .40 would adequately demonstrate the worst case scenario. The equations for calculating and applying the attenuation factor ( $\alpha$ ) are shown below. All of the equations are provided by Hunter and Schmidt (1990). In these equations,  $\underline{d}_0$  is the uncorrected study effect size,  $\underline{w}_i$  is the corrected weight for the study,  $\underline{v}_i$  is estimated study sampling error and  $\underline{D}_0$  is the uncorrected Ave( $\underline{d}$ ). Hunter and Schmidt (1990) explain that when unreliability is

present in the dependent measures, effect sizes are underestimated, sampling error increases (and can be estimated for each study) and the contribution of each study in the

$$\underline{a} = \sqrt{\underline{r}_{yy}}$$

$$\underline{d} = \underline{d}_o / \underline{a}$$

$$\underline{w}_i = \underline{N}_i \underline{a}_i^2$$

$$\underline{ve}_i = [(\underline{N}_i - 1) / (\underline{N}_i - 3)] [4 / \underline{N}_i] [1 + \underline{D}_o^2 / 8] / \underline{a}_i^2$$

$$Ave(\underline{d}) = \sum \underline{w}_i \underline{d}_i / \sum \underline{w}_i = \underline{D}$$

$$Var(\underline{d}) = \sum \underline{w}_i (\underline{d}_i - \underline{D})^2 / \sum \underline{w}_i$$

$$Var(\underline{e}) = \sum \underline{w}_i \underline{ve}_i / \sum \underline{w}_i$$

$$Var(\underline{\delta}) = Var(\underline{d}) - Var(\underline{e})$$

$$\underline{SD}_\delta = \sqrt{Var(\underline{\delta})}$$

final MA should be proportional to the reliability of the dependent measures in those studies. The individual study calculations are provided in Table 7.

Results of the MA based on the values in Table 7 are shown in Table 8. Ave(d) changed very little when the four studies were corrected for dependent measure reliability of .40. The attenuation in this instance had little impact because of the small study weightings assigned to the four corrected studies. Both Var(d) and Var(e) increased as anticipated and the relatively large magnitude of Var(e) again resulted in a negative value for Var(δ), SD<sub>δ</sub> = 0 and precluded the development of a credibility interval around δ.

Table 7. Study Values Adjusted for Reliability of the Dependent Measure

Study Reference	<u>N</u>	<u>d</u> <sub>0</sub>	<u>r</u> <sub>yy</sub>	<u>d</u>	<u>ve</u> <sub>i</sub>	<u>w</u> <sub>i</sub>
(Woodruff et al., 1976)	8	0.5425	.4	0.8579	1.7559	3.2
(Martin and Waag, 1978a)	8	0.2154	.4	0.3406	1.7559	3.2
(Martin and Waag, 1978b)	24	0.1242	.4	0.1965	0.4579	9.6
(Nataupsky, et al., 1979)	32	0.3120	.4	0.4934	0.3352	12.8
(Westra, 1982)	32	0.3476	1.0	0.3476	0.1336	32
(van der Pal, 1999)	12	0.0115	1.0	0.0115	0.4074	36
(Go et al., 2000)	36	-0.1462	1.0	-0.1462	0.1178	12

Table 8. Final Values for MA Corrected for Attenuation

Variables	Value
Ave( <u>d</u> )	0.17
Var( <u>d</u> )	0.0722
Var( <u>e</u> )	0.3064
Var( <u>δ</u> )=	-0.2341
<u>SD</u> <sub>δ</sub> =	0.0

## DISCUSSION

The results of this MA suggest a small, positive, performance benefit for pilot simulator training when that training includes simulator platform motion versus the same training without platform motion. And, although estimates of  $\text{Var}(e)$  can be overestimated when the analysis uses small sample sizes (Hunter & Schmidt, 1990), it appears that any variance across studies is due entirely to sampling error. In the current analysis, the estimate of  $\text{Var}(e)$  would indeed need to be a gross overestimate in order to conclude anything else because  $\text{Var}(e)$  is nearly five times the observed  $\text{Var}(d)$ .

This is a contradiction to the findings reported by Jacobs et al. (1990) that may well be due to the inclusion of the Ryan et al. (1978) data in their analysis. However, the results of this MA should not be taken as a resounding validation of the Pfeiffer and Horey (1987) work either. An effect size of  $d = 0.16$  is small at best and there are several reasons for being cautious in the interpretation of this number.

First, this study was based on a very small sample size. Considering the potential impact of including even one other study (e.g., Ryan et al., 1978) it should be clear that the paucity of data in this area is reason for concern. Another concern in the calculations is that homogeneity of variance was assumed because there were not data with which to determine otherwise. As Grissom and Kim (2001) suggest, using  $t$  and  $F$  from primary research (because these are commonly reported) indirectly assumes homoscedasticity because the use of  $t$  or  $F$  assumes so. Further, estimates of  $d$  can vary greatly in the presence of heteroscedasticity depending on which estimate of variance is used.

There are also a number of methodological issues within the available studies that call their inclusion here into question. These range from the addition of motion during

training for stall maneuvers to the no-motion condition by Woodruff et al. (1976) to the admission by Go et al. (2000) that their simulator may not have provided lateral acceleration cues appropriate for the maneuvers they tested during the training of the motion group.

There are certainly reasons that any true beneficial effect due to simulator motion during training would be small. Recall that one of the arguments in support of motion has relied on the theory of identical elements (Thorndike & Woodworth, 1901). The basic argument is that the greater the accuracy with which critical performance cues in the operational setting are replicated in the training setting, the better the skill transfer. MacKay (1982) presents an interesting caveat to this argument. In his addition to the theory, he suggests that prior experience with similar cues can strengthen linkages between those cues and subsequent responses that make learning the new task easier. This might be seen as training before the training in the current environment. By the time most student pilots enter flight training they have likely operated a variety of large moving vehicles including bicycles, riding lawn mowers, go-carts, cars, trucks and boats. They have spent their lives in a motion and gravity rich environment and they know how to interpret motion input via their visual and vestibular systems and respond accordingly. In this sense, it is not likely that the first motion cues they have to respond to are the ones they experience during flight training. It is just as likely that the new motion cues that will be encountered in the operational flight environment are the least likely to be simulated accurately. Finally, because of the prior high levels of experience with various motion rich environments, adaptation in responding to novel cues may be extremely rapid. This would explain why trainees who have apparently adopted inappropriate



control strategies when training without motion, as reported by Lee and Bussolari (1989) and Van der Pal (1999) are able to modify those strategies rapidly when provided with motion cues.

In most of the transfer studies cited in this paper, a rich visual environment was included for many if not all maneuvers evaluated. This is another reason that a beneficial influence of motion on training transfer may be minimized. Visual motion cues may well be strong enough to support the learning of most responses necessary to achieve proficiency. Even if visual cues alone do not overshadow the benefit of physical motion, visual cues in concert with feedback from cockpit instrumentation certainly might. How often are we really asking pilots to respond to a situation in which both visual motion cues and feedback from instruments provide inadequate cueing for proficient performance and, are those the only maneuvers for which motion is being advocated?

An even more relevant question might be how accurately can we measure performance in situations such as that described above? As indicated in Table 3, a variety of subjective assessment techniques were used in the studies included in this MA. How accurate are 4 or 12-point scales or hand scored data cards at capturing performance and discriminating among individuals in tasks with the characteristics of those described above or on any other task for that matter? Crosby and Parkinson (1979) demonstrated that measures of mental workload could discriminate between student pilots near the end of their training and experienced IPs when traditional, subjective ratings could not. Likewise, the workload measure they used (secondary task/memory search) allowed them to discriminate between students who differed in only 4 weeks of experience. They argue that mental workload measures may provide a more sensitive measure of pilot

proficiency. However, measures of mental workload are hardly the norm for assessing pilot performance in modern training programs or even in the existing transfer studies. Even evaluating performance in the simulators is problematic because, as Salas, Bowers and Rhodenizer (1998) pointed out, “often, high-fidelity simulators do not collect performance measures that can be readily used constructively in training evaluation” (p. 204). Boldovici (1992) also pointed to performance measurement as one of many reasons that there is a lack of evidence supporting motion. More specifically, he suggested that one focus of research should be the development of more reliable tools for assessing performance on unsafe tasks.

In the end, the question is not just whether there is an advantage to having motion but how valuable any existing advantage may be for pilot training? One should consider some of the costs associated with the addition of a motion platform. For trainees’, the increase in monetary costs can be substantial. Training time in a Level D (as defined in AC-120-40B) simulator typically costs between \$550 and \$1100 an hour. Limited availability (largely due to ownership costs) for certified simulators also means scheduling issues, travel costs and time away from the job for many trainees. For the owner/operators, Level D simulators can cost in the millions of dollars (although not all of this is attributable to the motion system). Motion platforms require more physical space, more computing power, greater environmental control, more manpower for support and result in higher maintenance costs.

In an attempt to extrapolate from any apparent positive effect of motion to the implications of that finding for pilot training, it should first be noted that the  $\delta$  of 0.16 in this case represents a 0.16 standard deviation in performance level. This is not directly

interpretable as either a savings in training time or a difference in the “safeness” of aircraft operation. Reasons for this are discussed and illustrated below.

Figures 1 and 2 illustrate two very different scenarios for hypothetical relationships between potential learning/performance curves of pilots training with or without motion and potential criterion levels of performance (lines A and B, Y-axis) across arbitrary units of training (X-axis). Figure 1 suggests that, during training, trainees receiving no motion will never achieve the same level of performance as those receiving motion. This can then be interpreted in two ways depending on which criterion level of performance is assumed. If the criterion level of performance is set at A, only pilots training in the presence of motion cues will ever be able to reach proficiency in the simulator. The 0.16 standard deviation difference between the groups will thus require that the no-motion trainees receive additional training in the real aircraft to close the gap. While we have no idea how much aircraft training will be required to close the performance gap, based on the evidence summarized in this report it is likely that any differences will disappear during or just after the first training trial in the aircraft.

If, on the other hand, the criterion for demonstrating proficiency is set at B, both groups will meet the criterion during training with a time savings for the motion group equal to  $t$ . However, determining  $t$  is no simple matter because we do not know what the learning curves under the conditions of motion and no motion really look like. The savings could be less time than it takes to fly a single maneuver or it could include hundreds of trials. As indicated earlier, our relatively insensitive measures of pilot performance would make the development of such learning curves problematic and the shape of the curve would most certainly depend on the tasks. Again, the question of

performance measurement becomes relevant. If our performance measures lack discriminatory power, we will not be able to assess the value of simulator platform motion in terms of either monetary cost or safety.

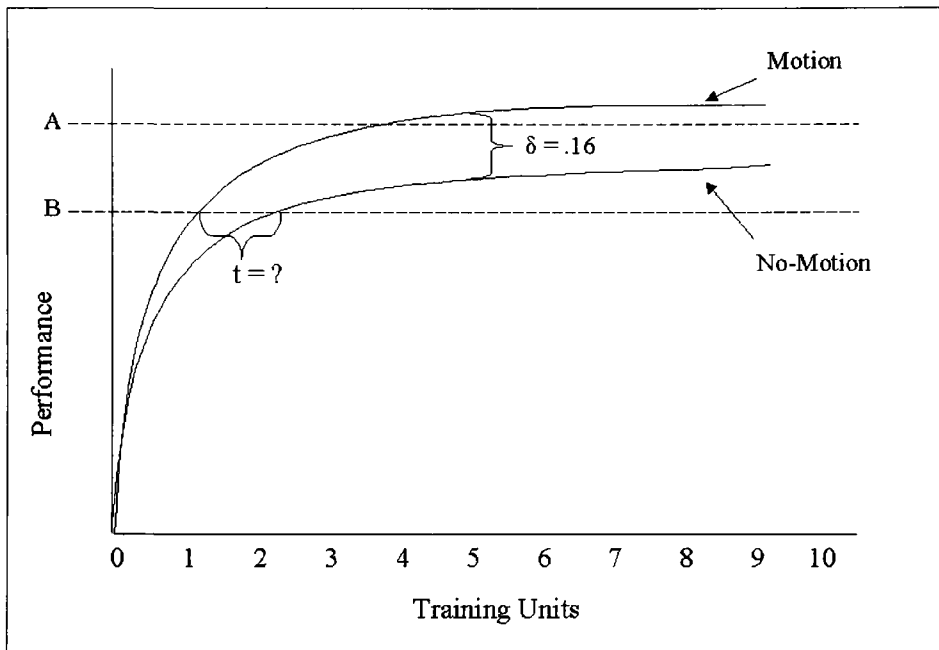


Figure 1. Hypothetical Learning Curves and Pilot Performance Criteria for Motion and No-Motion in Simulator Training – Scenario 1.

Figure 2 shows a slight modification to the scenario in Figure 1. Here, motion is beneficial early in training but the advantage disappears with further simulator time. Keeping in mind that the chart is only hypothetical, it is possible that we manage to overlook a larger benefit for motion early in training. Does the additional no-motion training remain cost effective? We cannot answer this question without more knowledge about the learning curves associated with specific tasks for both motion and no-motion trained pilots. And again, performance measurement will likely be an issue.

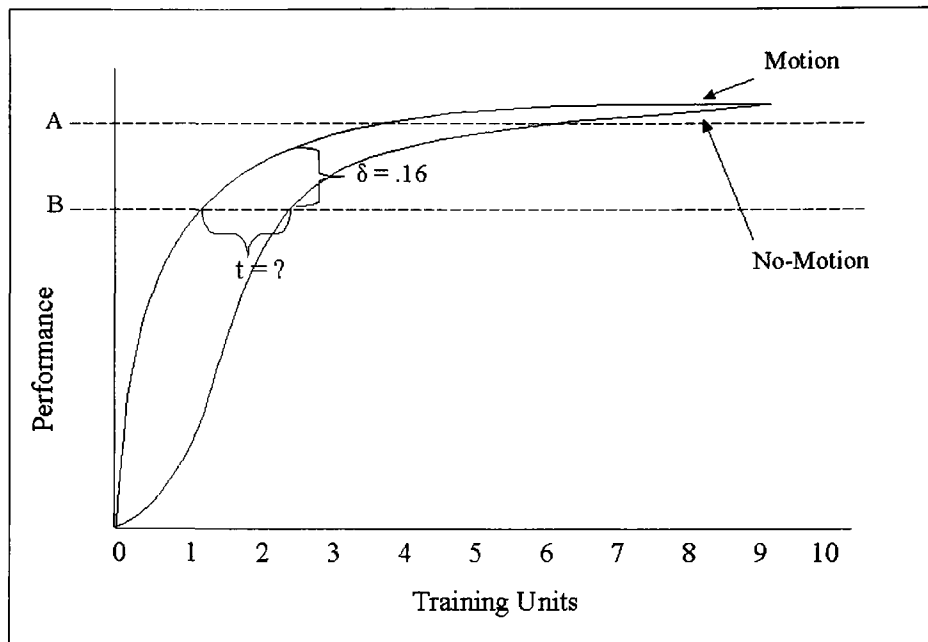


Figure 2. Hypothetical Learning Curves and Pilot Performance Criteria for Motion and No-Motion in Simulator Training – Scenario 2.

## CONCLUSION

The dearth of empirical studies on this topic is somewhat alarming considering the overall amount of conjecturing that has gone over the past 40 years and the seeming importance of the topic. There are a variety of reasons for this. Access to the equipment is limited and expensive (Salas et al., 1998), research participants are typically limited to pilots in on-going training programs and attempting to conduct such research in operational settings poses an entire host of problems. The research may be intrusive to the training environment. The researchers may be dependent on personnel who are less motivated when it comes to conducting a well controlled experiment (i.e., IP's). Curriculum limitations may influence the ability to control participants, scheduling, selection of maneuvers, and general data collection. Boldovici (1992) argued that true

transfer studies will never answer the question of whether or not platform motion is needed anyway. He suggested that the maneuvers that most people believe to be impacted by motion can not be tested in the real aircraft. This lends relevance to the quasi-transfer design but few of these studies have been conducted.

It seems that the opinions of engineers and researchers are as divergent as ever. The suggestion by Buckhout et al. (1963) that “blind dedication to the achievement of realism of simulation can sometimes frustrate the whole intent of the research effort” (p. 41) has been echoed through four decades of technological development during which time simulators have changed substantially while training programs and performance measurement systems have not (Salas et al. 1998).

In summation, the results of this MA provide some evidence of a slight but positive effect of simulator platform motion on transfer of pilot training. Several factors have been discussed that may mitigate the apparent effect of motion. These include prior awareness of motion cues, cue redundancy made available by visual motion cues or instrument feedback, the type of maneuver being trained and relative insensitivity in the performance measurement tools used to detect differences between pilots trained either with or without simulator motion. Potential scenarios have also been presented for assessing the value of the estimated true effect size for motion. The task of identifying the particular scenario on which to base a final value calculation would also benefit from a more accurate system of performance measurement.

## REFERENCES

- Anderson, J.R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Boldovici, J.A. (1992). *Simulator Motion*. U.S. Army Research Institute for the Behavioral and Social Sciences. ARI Technical Report 961.
- Buckhout, R., Sherman, H., Goldsmith, C.T. and Vitale, P.A. (1963). *The Effect of Variations in Motion Fidelity During Training on Simulated Low-Altitude Flight*. 6570<sup>th</sup> Aerospace Medical Research Laboratory, Wright-Patterson, AFB, OH. AMRL-TDR-63-108.
- Bürki-Cohen, J., Go, T.H., & Longbridge, T. (2001) Flight simulator fidelity considerations for total air line pilot training and evaluation. Paper presented at the AIAA Modeling and Simulation Technologies Conference, Montreal, Canada. (AIAA Paper 2001-4425).
- Bürki-Cohen, J., Soja, N.N., & Longbridge, T. (1998). Simulator platform motion – the need revisited. *The International Journal of Aviation Psychology*, 8(3), 293-317.
- Crosby, J.J. and Parkinson, S.R. (1979). A dual task investigation of pilots' skill level. *Ergonomics*, 22(12), 1301-1313.
- Glass, G.V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.
- Go, T.H., Buerki-Cohen, J. and Soja, N.N. (2000) The effect of simulator motion on pilot training and evaluation. Paper presented at the AIAA Modeling and Simulation Technologies Conference, Denver, CO. (AIAA Paper 2000-4296).

- Grissom, R.J. and Kim, J.J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6(2), 135-146.
- Hall, J.R. (1978). Motion versus visual cues in piloted flight simulation (Rep. No. AGARD-CP-249). Paper presented at the Flight Mechanics Panel Specialist Meeting on Piloted Aircraft Environment Simulation Techniques, Brussels, Belgium.
- Hedges, L.V., & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin*, 88, 359-369.
- Holding, D.H. (1976). An approximate transfer surface. *Journal of Motor Behavior*, 8, 1-9.
- Holt, R.W., Hansberger, J.T. and Boehm-Davis, D.A. (2002). Improving rater calibration in aviation: A case study. *The International Journal of Aviation Psychology*, 12(3), 305-330.
- Hopkins, C.O. (1975). How much should you pay for that box? *Human Factors*, 17(6), 533-541.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newbury Park, CA: SAGE Publications, Inc.
- Jacobs, R.S., Prince, C., Hays, R.T. and Salas, E. (1990). *Meta-Analysis of the Flight Simulator Training Research*. Naval Training Systems Center, Orlando. NTSC-89-006.
- Jacobs, R.S. and Roscoe, S.N. (1975). Simulator cockpit motion and the transfer of initial flight training. In *Proceedings of Human Factors Society Annual Meeting*, Dallas, TX. p. 218-226.
- Jones, S. and Franklin, B.D. (1999). *Determining motion-cueing requirements for the*



- Advanced Amphibious Assault Vehicle (AAAV) Driver Simulator*. Naval Air Warfare Center Training Systems Division. Orlando, FL. Marine Corps Report 99-03.
- Koonce, J.M. (1979). Predictive validity of flight simulators as a function of simulator motion. *Human Factors*, 21(2), 215-223.
- Lee, A.T. and Bussolari, S.R. (1989). Flight simulator platform motion and air transport pilot training. *Aviation, Space, and Environmental Medicine*, 60(2), 136-140.
- MacKay, D.G. (1982). The problems of flexibility, fluency, and speed/accuracy trade-off in skilled behavior. *Psychological Review*, 89, 483-506.
- Martin, E.L. and Waag, W.L. (1978a). *Contributions of platform motion to simulator training effectiveness: Study I – Basic contact*. AFHRL-TR-78-15.
- Martin, E.L. and Waag, W.L. (1978b). *Contributions of platform motion to simulator training effectiveness: Study II – Aerobatics*. AFHRL-TR-78-52.
- McDaniel, W.C., Scott, P.G., & Browning, R.F. (1983). *Contribution of platform motion simulation in SH-3 helicopter pilot training*. US Navy, Training Analysis and Evaluation Group. Orlando, FL.
- Nataupsky, M., Waag, W.L., Weyer, D.C., McFadden, R.W. and McDowell, E. (1979). *Platform motion contributions to simulator training effectiveness: Study III. Interaction of motion with fields of view*. AFHRL-TR-79-25.
- Osgood, C.E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, 56, 132-143.
- Pfeiffer, M.G. and Horey, J.D. (1987). *Training effectiveness of aviation motion simulation: A review and analysis of the literature*. Special Report 87-007. Naval Training Systems Center, Department of the Navy, Orlando, FL.

- Ryan, L.E., Scott, P.G. and Browing, R.F. (1978). The effects of simulator landing practice and the contribution of motion simulation to P-3 pilot training. TAEG-63; ADA0611434.
- Salas, E., Bowers, C.A. and Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *International Journal of Aviation Psychology*, 8(3), 197-208.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Slavin, R.E. (1986). Meta-analysis in education: How has it been used? *The Educational Researcher*, 3(18), 6-15.
- Strachan, I.W. (1997). To move or not to move? That is the question. In the *Proceedings of the 19<sup>th</sup> Interservice/Industry, Training, Simulation and Education Conference*. Orlando, FL: National Training Systems Association.
- Szczepanski, C., & Leland, D. (2000). *Move or not to move? Continuous question*. Paper presented at the AIAA Modeling and Simulation Technologies Conference, Denver, CO. (AIAA No. 2000-4297).
- Thorndike, E.L. and Woodworth, R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247-261.
- Van der Pal, J. (1999). The effect of simulator motion on parameter training for F-16 pilots. *Engineering psychology and cognitive ergonomics. Vol. 3 – Transportation systems, medical ergonomics and training*. In Proceedings of the 1st International

Conference on Engineering Psychology and Cognitive Ergonomics, Oxford, United Kingdom, Oct. 1998, 267-275.

Westra, D.P. (1982). *Simulator design features for carrier landing. II. In-simulator transfer of training*. Canyon Research Group, Inc. CRG-TR-82-011; NAVTRAEQUIPC-81-0105-1.

Woodruff, R.R., Smith, J.F., Fuller, J.R. and Weyer, D.C. (1976). Full mission simulation in under-graduate pilot training: An exploratory study. AFHRL-TR-76-84.

**APPENDIX**  
**STUDY SUMMARIES**

1. Authors: Buckhout, Sherman, Goldsmith and Vitale
2. Date of Pub: 1963
3. Transfer or Quasi-Transfer: Quasi Transfer
4. Participants in Exp Group: 8
5. Participants in Control Group: 8
6. Participant Experience Level: low hours.
7. Simulator Type Used: Grumman Multipurpose Motion Simulator
8. Degrees of Freedom for Motion: 3
9. FOV: 4 inch CRT
10. Training Type: Blocked – 15 training trials and 3 transfer trials.
11. Maneuvers assessed: Tracking task during low altitude flight.
12. Dependent Measures: Objective measures of RMSE, time on target, violations of altitude limit and crashes
13. Data collection technique: Electronic from the sim.
14. Analysis type: ANOVA.
15. MA data available: Overall F but it was for 8 groups of varied levels of motion. No good data for the MA.

1. Authors: Jacobs and Roscoe
2. Date of Pub: 1975
3. Transfer or Quasi-Transfer: Transfer
4. Participants in Exp Group: 9
5. Participants in Control Group: 9
6. Participant Experience Level: undergraduate trainees.
7. Simulator Type Used: Singer-Link GAT-2
8. Degrees of Freedom for Motion: not stated
9. FOV: not stated
10. Training Type: Blocked – does not specify trial number per maneuver.
11. Maneuvers assessed: 11 but does not identify them.
12. Dependent Measures: Subjective – time to criterion, trials to criterion and error (violations of limits set by private pilot flight test).
13. Data collection technique: Paper/pencil for IP scores.
14. Analysis type: Covariance on a variety of scores.
15. MA data available: None – could not use study. No reliable differences were reported between motion and no-motion groups during transfer. Study employed a random washout motion condition which was interesting – no subjects in that condition reported noticing random reversal of bank motion during the training.

1. Authors: Woodruff, Smith, Fuller and Weyer
2. Date of Pub: 1976
3. Transfer or Quasi-Transfer: Transfer
4. Participants in Exp Group: 4
5. Participants in Control Group: 4
6. Participant Experience Level: Less than 50 hours
7. Simulator Type Used: Advanced Simulator for Undergraduate Pilot Training (ASUPT)
8. Degrees of Freedom for Motion: 6
9. FOV: not reported
10. Training Type: Proficiency based – followed standard training syllabus based on IP ratings
11. Maneuvers assessed: All in program collapsed into Basic and Presolo, Advance Contact, Instruments, Formation and Navigation
12. Dependent Measures: Subjective
13. Data collection technique: hours to criterion for the five training segments above based on IP ratings – no mention of actual rating technique
14. Analysis type: ratio of hours needed for the two groups (hours E : hours C) Raw data reported (average hours per participant per training segment)
15. MA data available: used raw hours per participant per segment to calculate two-tailed t-test. Calculated a study  $d$  based on five t-test results.

1. Authors: Ryan, Scott and Browning
2. Date of Pub: 1978
3. Transfer or Quasi-Transfer: transfer to P-3
4. Participants in Exp Group: 39
5. Participants in Control Group: 11
6. Participant Experience Level: relatively low – completing undergrad curriculum
7. Simulator Type Used: 2F87F – P-3 Orion – four engine turbo-prop
8. Degrees of Freedom for Motion: 6
9. FOV: 50 horizontal and 38 vertical
10. Training Type: proficiency on 5 main tasks
11. Maneuvers assessed: Abort Four Engine, Abort Three Engine, Engine Failure After Refusal, Instrument Tasks (Holding, Non-precision Approach TACAN, VOR, NDB, LOC, Precision Approach GCA, ILS, Instrument Procedures), Landings (Normal Landings, Approach Flap Landings, Three Engine Landings)
12. Dependent Measures: Subjective – UBAA determined aircraft trials to proficiency.
13. Data collection technique: UBAA hand scored w/ paper and pencil
14. Analysis type: repeated measures  $F$  – across maneuver.
15. MA data available: repeated measures  $F(1,48) = 3.21, p = .079$ . The Jacobs, Prince, Hays and Salas MA reports a point biserial correlation of  $-.297$  with  $N=50$



1. Authors: Martin and Waag
2. Date of Pub: 1978a
3. Transfer or Quasi-Transfer: Transfer
4. Participants in Exp Group: 8
5. Participants in Control Group: 8
6. Participant Experience Level: undergraduate trainees – average flight experience = 28.8 hours.
7. Simulator Type Used: Advanced Simulator for Pilot Training (ASPT)
8. Degrees of Freedom for Motion: 6
9. FOV: “full” but specifics not given
10. Training Type: Blocked – 10 sorties in the ASPT
11. Maneuvers assessed: Three categories of sortie – Basic Work (12 maneuvers), Pattern Work (4 maneuvers), Mission Profiles (all 16 prior maneuvers)
12. Dependent Measures: Subjective – IP Ratings used, Task frequency – repetition to solo.
13. Data collection technique: IP ratings for two evaluation flights in T-37 were collected in log books – based on 12 point scale – 1-3=unsat, 4-6=fair, 7-9=good, 10-12=excellent. (short-term measure of transfer)  
  
IP ratings across Task Frequency up to solo also recorded for 8 maneuvers – Takeoff, Straight-in Approach, Landing, Overhead Pattern, Overhead Landing, Slow Flight, Power-On Stall and Traffic Pattern Stall. An average rating per student per maneuver was calculated. (Long-term measure of transfer)
14. Analysis type: sixteen split plot ANOVAs performed

15. MA data available: F-values from 16 ANOVA's for the maneuvers assessed during two transfer flights in the T-37 and a priori t-tests for Task Frequency data.

1. Authors: Martin and Waag
2. Date of Pub: 1978b
3. Transfer or Quasi-Transfer: Transfer
4. Participants in Exp Group: 12
5. Participants in Control Group: 12
6. Participant Experience Level: undergraduate trainees – average flight experience = 28.8 hours.
7. Simulator Type Used: Advanced Simulator for Pilot Training (ASPT)
8. Degrees of Freedom for Motion: 6
9. FOV: “full” but specifics not given
10. Training Type: Blocked – 5 sorties in the ASPT for basic aerobatics then transfer, then two sorties in the ASPT for advanced aerobatics
11. Maneuvers assessed: Basic Aerobatics = aileron roll, split s, loop and lazy 8.  
Advanced aerobatics = Immelmann, barrel roll, cuban 8, and clover leaf.
12. Dependent Measures: Subjective – IP scoring done on special data cards for entry airspeed, bank at entry, pitch rate control, ground track control, etc. In total, 40 measures were taken across the 8 maneuvers. Evals per maneuver ranged from 3 (Aileron Roll) to 7 (Lazy 8).
13. Data collection technique: special data cards used for each maneuver and averaged across transfer trials (these varied in number).
14. Analysis type: 40 univariate  $F$  tests and a priori t-test reported on same means.
15. MA data available:  $F$ -values from 40 ANOVA's and t-tests for the maneuvers assessed during the transfer flights in the T-37.

1. Authors: Nataupsky, Waag, Weyer, McFadden and McDowell
2. Date of Pub: 1979
3. Transfer or Quasi-Transfer: Transfer
4. Participants in Exp Group: 16
5. Participants in Control Group: 16
6. Participant Experience Level: Undergraduates transitioning to the T-37 – 25 to 64 hours of flight experience.
7. Simulator Type Used: Advanced Simulator for Pilot Training (ASPT)
8. Degrees of Freedom for Motion: 6
9. FOV: 300 Horizontal x 150 Vertical (a second FOV treatment level used 48 Horizontal and 36 Vertical)
10. Training Type: blocked – 4 trials in sim then one transfer trial
11. Maneuvers assessed: Takeoff, Steep Turn, Slow Flight, Straight-In (before Glidepath), Straight-In (On Glidepath)
12. Dependent Measures: IP eval on 8 point rating scale and values recorded with special recording cards as follows:
  - Take-off: Pitch Range, Rotation Speed, Ground Deviation, Liftoff, IP Rating
  - Steep turn: Altitude Range, Bank Range, Airspeed Range, IP Rating
  - Slow Flight: Altitude Range, Airspeed Range, Heading Range, IP Rating
  - Straight-In (Before Glidepath): Altitude Range, Airspeed Range, Centerline deviation
  - Straight-In (On Glidepath): Altitude Range, Airspeed Range, Centerline deviation, IP Rating

13. Data collection technique: special rating cards used by IP's
14. Analysis type: two-factor ANOVA for each measure
15. MA data available:  $F$  for every measure listed above (from table 9 on page 14) is given below:

1. Authors: Westra
2. Date of Pub: 1982
3. Transfer or Quasi-Transfer: quasi transfer
4. Participants in Exp Group: 16
5. Participants in Control Group: 16
6. Participant Experience Level: mixed but no carrier landing experience.
7. Simulator Type Used: Visual Technology Research Simulator (VTRS) – T-2C jet
8. Degrees of Freedom for Motion: 6
9. FOV: 160 horizontal and 80 vertical – was manipulated as second factor in study
10. Training Type: Blocked – 40 training trials, 16 transfer trials.
11. Maneuvers assessed: Circling approach and landing (on simulated Carrier).
12. Dependent Measures: Objective - touchdown wire accuracy, glideslope tracking, lineup tracking, angle of attack.
13. Data collection technique: electronic info from simulator.
14. Analysis type: ANOVA.
15. MA data available: 4 ANOVAs reported – note these are as if repeated measures as scores during transfer were collapsed across two, 8 trial blocks.

1. Authors: Lee and Bussolari
2. Date of Pub: 1989
3. Transfer or Quasi-Transfer: Neither – no training was done – only testing
4. Participants in Exp Group: 6 Exp. 1, 8 Exp. 2
5. Participants in Control Group: 6 Exp. 1, 8 Exp. 2
6. Participant Experience Level: 2.4 yr. Average in Exp 1, no hours in AC model in Exp. 2.
7. Simulator Type Used: Boeing 727-700
8. Degrees of Freedom for Motion: 6
9. FOV: did not report
10. Training Type: No training.
11. Maneuvers assessed: 3 scenarios, variety of maneuvers.
12. Dependent Measures: Subjective and Objective
13. Data collection technique: Rating scales on paper and Electronic from the sim.
14. Analysis type: ANOVAs.
15. MA data available: Few numbers given – mostly just general statements about lack of difference between groups.

1. Authors: van der Pal
2. Date of Pub: 1999
3. Transfer or Quasi-Transfer: quasi transfer
4. Participants in Exp Group: 6
5. Participants in Control Group: 6
6. Participant Experience Level: high – ex-F-16 pilots - retired
7. Simulator Type Used: Re-configurable – F16
8. Degrees of Freedom for Motion: 6
9. FOV: 142 horizontal and 110 vertical
10. Training Type: blocked – aerobatics – 20 trials (not reported); weapon delivery maneuver – 12 trials
11. Maneuvers assessed: weapons delivery
12. Dependent Measures: Objective – data from sim
13. Data collection technique: electronic from sim
14. Analysis type: ANOVA on various performance parameters
15. MA data available:  $F(1,10) = 1.22$ ,  $p=.3$  for absolute dev. Altitude at apex of maneuver – graphical data shows motion condition with larger error here.  
 $F(1,10)=1.13$ ,  $p=.31$  for roll correction frequency band width – in favor of motion group.



1. Authors: Go, Burki-Cohen and Soja
2. Date of Pub: 2000
3. Transfer or Quasi-Transfer: Quasi-Transfer
4. Participants in Exp Group: motion – 18/16 depending on DV assessed
5. Participants in Control Group: no motion – 18/19 depending on DV assessed
6. Participant Experience Level: high – regional airline pilots during recurrent training
7. Simulator Type Used: FAA qualified level C – 30 passenger, turbo prop, twin engines
8. Degrees of Freedom for Motion: 6
9. FOV: 150 horizontal x 40 vertical to each pilot
10. Training Type: Criterion based – within framework of ongoing training program
11. Maneuvers assessed: Engine failure on either rejected take-off (RTO) or continued take-off (V1 cut)
12. Dependent Measures: Both Subjective and Objective but only objective reported.
13. Data collection technique: electronic via sim.
14. Analysis type: t-tests
15. MA data available: p-values and sample sizes for 6 measures of transfer – all for V1-cut maneuver only – no differences reported for RTO maneuver.
  - a. Integrate Airspeed Exceedance (ne=18) (nc=19),  $p=.006$  – extrapolated  $t(35)=2.65$ (one-tailed).
  - b. STD Pitch Angle (ne=18) (nc=19),  $p=.025$  – extrapolated  $t(35)=-2.03$ (one-tailed).
  - c. Wheel Reversals (ne=18) (nc=19),  $p=.059$  – extrapolated  $t(35)=1.60$ (one-tailed).
  - d. Pedal Reversals (ne=18) (nc=19),  $p=.008$  – extrapolated  $t(35)=-2.53$ (one-tailed).

- e. Integrated Yaw Activity (ne=16)(nc=18),  $p=.024$  – extrapolated  $t(32)=-2.06$ (one-tailed).
- f. RMS Heading Deviation(ne=16)(nc=18),  $p=.354$  – extrapolated  $t(32)=-.38$ (one-tailed).