Exploring Learning Measures During Training on a Truck-Dispatcher Task

We used an inter-disciplinary approach to assess the validity of using alternative behavioral variables including task speed, information seeking, and types of errors as operational definitions of learning rather than relying on performance scores alone. Our study was a re-evaluation of a sample from a larger data set (Palumbo, 2008). Undergraduate students (N = 145) participated in a moderately difficult truck-dispatcher task (Steele-Johnson & Perlow, 1989) in which they received, processed, and shipped orders of military supplies. We categorized participants into greatest-learning versus least-learning groups based on the amount of change in their performance scores, and then compared task speed, information seeking, and types of errors between the two groups. Results indicated that task action speed and patterns of errors could be used as learning indicators for this task. Results also provided initial evidence of the efficacy of using alternative measures other than final performance score to assess learning during training.

Organizations have highlighted the importance of and spent substantial sums on training (Cascio & Aguinis, 2005; Facteau, Dobbins, Russell, Ladd, & Kudisch, 1995). Workplace training can help employees adapt quickly to required changes in work tasks and improve their productivity (Quiñones, 1997). Thus, the focus of researchers and practitioners alike has been on ways of improving training and training outcomes. Certainly cognitive ability plays a large part in training outcomes (e.g., Hunter & Schmidt, 1996; Schmidt & Hunter, 1998), but research has shown that there are other factors, such as training type (Baldwin & Ford, 1998; Keith & Frese, 2008) or trainee characteristics (Warr & Bunce, 1995) that also influence learning during training. Researchers have operationalized learning outcomes in workplace training generally as scores on paper-and-pencil or performance tests (Arthur, Bennett, Edens & Bell, 2003). Scores from paper-and-pencil or performance tests are important because they can be directly related to the company bottom line. However, performance scores might not tell us enough. For example, it would be useful to know what was learned and when.

There is certainly precedent in the training literature for examining measures of learning other than performance scores. For example, Kraiger, Ford and Salas (1993) called for multiple outcome measures, including skill-based, cognitive, and affective outcomes, in order to more thoroughly evaluate training effectiveness. Also, research (Austin & Bobko, 1985; Erez, 1990) has indicated there is some value in distinguishing between performance quantity outcomes (i.e., speed) and performance quality outcomes (i.e., accuracy). Moreover, when designing computerized training systems, such as intelligent tutoring systems which adapt training based on trainee errors, researchers must have access to indicators of learning that occur well before the final performance score (Steele-Johnson & Hyde, 1997). Identifying behaviors indicative of learning and then using these behaviors as signals during training to tailor training methods to the individual could increase the effectiveness of training.

When evaluating the effectiveness of workplace training, a central question is whether the training results in learning (e.g., Kraiger & Jung, 1997). However, merely calculating performance scores, as most training evaluations do, may provide an incomplete assessment of trainee learning. More specifically, it might be useful to also identify what was learned, that is, the content of learning. For example, individuals who make different types of errors might obtain similar performance scores, but trainers could be more effective if they knew specifically what kinds of errors trainees made. Moreover, in developmental learning research, in addition to assessing participants’ end results, such as performance scores, researchers also determine

* Faculty mentor
Trainees’ completion of initial task actions may be a distinguishing variable among participants’ improvement in time taken to complete initial task actions, relative to those in the least-learning group. We hypothesized that with task practice, trainees or training systems could make to enhance trainee learning. We sought to determine alternative operational definitions of learning (indicator variables) that occur during training. These indicator variables could be used to differentiate early in the training process between those who are learning the material and those who may not be learning as quickly in order to modify training parameters and improve training success. We tested the validity of using several different alternative variables as indicators of learning, including task speed, information seeking, and types of errors or mistakes.

**Task Speed**

**Hypothesis 1:** Trainees’ completion of initial task actions will occur more quickly with repeated task practice. As participants gain experience and task knowledge through repeated task practice, it is assumed that they will demonstrate improved skill development as evidenced by faster performance (e.g., Kraiger et al., 1993). Therefore, one factor that might be indicative of task learning is an increase in speed of task actions. Specifically, we hypothesized that with task practice participants would be more prepared to begin the task and therefore better able to complete initial task actions quickly, resulting in increased productivity and overall performance scores.

**Hypothesis 2:** Individuals in the greatest-learning group will demonstrate greater improvement in speed of initial task actions, relative to those in the least-learning group. Overall we expected most participants to display some improvement in the time taken to complete initial task actions after repeated practice (Hypothesis 1); however, the difference between participants’ improvement in time taken to complete initial task actions may be a distinguishing variable indicative of task learning. Comparing the improvement in the time participants take to complete an essential task action is a measure of learning based on improvement in speed that is often used in developmental and educational research (Gettinger & White, 1979; Ivkovich et al., 2000). However, researchers have rarely used this measure as an indicator of learning in I/O research (see Singer & Gaines, 1975, for an exception). For comparison purposes we operationally defined participants whose performance scores displayed the greatest increase as the greatest-learning group and those whose performance scores displayed the least increase as the least-learning group, and we omitted those individuals whose performance scores displayed moderate increases. We not only expected to see differences in the time individual participants took to complete initial task actions, we also hypothesized that participants whose performance scores showed the most improvement would demonstrate the most increase in speed of initial task actions.

**Hypothesis 3:** Amount of initial information seeking will be positively related to final performance score. For decades, researchers have viewed feedback as essential for learning and performance (Ilgen, Fisher & Taylor, 1979). Ashford and Cummings (1983) proposed that individuals engage in feedback-seeking behavior in order to reduce uncertainty about what goals to pursue. Participants who initially seek out information regarding task instructions and rules might be more likely to learn a task than those who do not seek out this information. Thus, another factor that may distinguish participants who display the greatest learning might be information-seeking behavior occurring during the beginning of training. The amount of initial information seeking should be positively related to final performance scores.

**Hypothesis 4:** Number of initial mistakes will be positively related to final performance scores. Traditionally, researchers have associated errors made during training with punishment (Skinner, 1953) and viewed errors as detrimental to learning (Bandura, 1986). However, more recent research (e.g., Dormann & Frese, 1994) has indicated that errors may be beneficial to learning in that they actually boost exploratory behavior, which in turn leads to strategy refinement and development. Researchers have shown that training which presents errors as useful or encourages errors leads to better performance on transfer or novel tasks than training which presents errors as dysfunctional or instructs learners to avoid errors (e.g., Keith & Frese, 2008). Additionally, early mistakes during training might be reflective of a more exploratory initial learning strategy. The recognition that one has made a mistake can cause further exploratory behavior.
Based on previous research demonstrating the beneficial effects of making mistakes during training (e.g., Lorenzet, Salas & Tannenbaum, 2005), we expected that trainees who made a greater number of errors early in training would gain more experience with the task and would perform better than those who did not experience these errors early on.

**Hypothesis 5:** Individuals in the greatest-learning group will display different types of errors relative to those in the least-learning group. Researchers in the training and motivation literature have demonstrated repeatedly that performance quantity and performance quality are not necessarily positively related (e.g., Erez & Arad, 1986). For instance, researchers have demonstrated that there can be a speed/accuracy trade-off in which increased speed of performance does not necessarily improve accuracy of performance, and vice versa (e.g., Stanton & Julian, 2002). Therefore, it may be beneficial for researchers interested in training to assess not only trainees’ improvement in task speed (Hypotheses 1 and 2) but also the types of errors trainees make. Types of errors might be another factor that distinguishes participants who have achieved the greatest learning from those who have learned the least.

**Hypothesis 6:** Individuals in the greatest-learning group will display different patterns of change in errors during training, compared to those in the least-learning group. Different error patterns could have different effects on performance, and thus individuals demonstrate their learning through changes in the types of errors they make with task practice. This could happen in two ways. Participants who have made certain errors early in task practice may learn from those errors, make them less frequently, and go on to display improved performance. Indeed, van der Linden, Sonnentag, Frese and van Dyck (2001) observed that initial errors can sometimes cause improvements in task strategy use. On the other hand, some errors might enhance performance. For instance, ignoring certain rules in a task in favor of other more essential ones might be a task strategy used by those who have better-learned the material. We expect those in the greatest-learning group will display different patterns of changes in errors, relative to those in the least-learning group.

**Method**

**Participants**
The current study examines part of a larger data set $N = 246$ (Palumbo, 2008). The subset of the data included in the analyses described here were from 145 undergraduate students (31 men and 114 women) from a medium-sized, midwestern university. Participants ranged in age from 18 to 53 years with a mean age of 21.72 years, $SD = 4.97$. Participants had received extra credit points that they could apply toward their course grade.

**Task Description and Procedure**
The larger study obtained a data set using an adaptation (Palumbo, 2008) of a well-studied, moderately difficult, truck-dispatcher task (Steele-Johnson & Perlow, 1989). All participants received task instructions and then completed five 10-min trials of the task. The task was an MS-DOS based computer simulation of duties required of a truck-dispatcher. Participants were required to receive, process, and ship orders of military parts and supplies to three areas within an Area of Responsibility. Running task time was defined as the time spanning one work week. Thus the task time began at 9:00 a.m. on Monday, and ended at 5:00 p.m. on Friday. Each hour interval in the task was equivalent to 15 s in real time, and each day in the task was equivalent to 2 min in real time. Prior to the task, participants read task instructions including seven rules concerning truck-capacity restrictions, time schedules, and delivery area restrictions.

Prior research has demonstrated that task instructions emphasizing either performance quality or performance quantity requirements can influence the speed/accuracy trade-off that is sometimes observed in individual performance (e.g., Stanton & Julian, 2002). Task instructions in the current study did not specifically stress either performance quantity (speed of shipment) or performance quality (obeying task rules); therefore, the task instructions should not have influenced one outcome more than another. Experimenters told participants that the purpose of the task was to accept incoming orders for processing, process those orders appropriately, and dispatch trucks quickly and correctly. Experimenters told participants also that they would receive points for each unit shipped and that points would be subtracted when rules were not followed.

In the truck-dispatcher task, subjects were required to ship items displayed in the service window either onto a truck or through a pick-up window (depending on item code). There were seven rules constraining how participants were to complete task actions. For instance, participants were required to limit loads on a truck to a certain level, send a truck to no more than three out of five shipment zones, and ship orders as indicated either onto a truck or to a pick-up window. See appendix for information on specific rules.

**Learner Type**
We created two groups of participants: those who demonstrated the greatest learning versus the least learning. This was a multi-step process. First, we computed a performance difference score by subtracting the Trial 1 score from the Trial 5 score for each par-
participant. Then we ordered the participants from those showing the greatest increases in performance scores from Trial 1 to Trial 5 to those showing the smallest increases (and including decreases) from Trial 1 to Trial 5. Finally, we identified as the greatest-learning group the 33% of participants showing the greatest increase in performance scores, and the least-learning group as the 33% of participants showing the smallest increase (including decreases). Final performance scores varied widely in both groups. There were 49 participants in the greatest-learning group (40 women and 9 men) with a mean age of 21.70 years, SD = 5.90 and 48 participants in the least-learning group (36 women and 12 men) with a mean age of 20.82 years, SD = 3.22. We omitted the 48 participants showing moderate increases (the middle 33%).

Composite Task Performance
We operationally defined composite task performance as participants’ performance scores in each trial. Participants received 5 points for each unit of office equipment correctly shipped and lost 10 points for each rule violation.

Subtask Performance
We operationally defined the subtask performance as the number of times each task action was completed during a trial. These task actions are listed below. Although not all actions resulted in points added or subtracted from the performance score, all actions required time to complete, which ultimately affected performance score.

Shipped truck. When participants dispatched one of the three trucks, and the truck was permitted to leave (did not exceed truck-capacity by over 20%), the computer software recorded the truck shipment and time associated with it. We operationally defined the subtask performance for this action as the time (seconds) taken to dispatch the first truck during a task trial.

Sent out pick-up. The computer software recorded the time at which participants placed pick-up orders in the pick-up window. We operationally defined the subtask performance for this action as the time (seconds) participants took to dispatch the first pick-up order during a trial.

Rule violations. The computer software recorded each time a participant violated one of the seven rules (see Appendix). We operationally defined the subtask performance for this action as the number of rule violations (both for individual rule violations and for total number of violations) made during a trial.

Viewing of Rules 1 through 7. At any time during the task, participants could view a specific rule on the computer screen by pressing the corresponding number 1 through 7 on the keyboard. We operationally defined subtask performance for this action as the number of rules viewed during a trial.

Results
Task Action Speed as a Measure of Learning (Hypotheses 1 and 2)
To analyze the effects of task practice on participants’ speed of task actions, we conducted a repeated measures analysis examining the effect of practice on time to complete initial task actions. Because we posited that practice would affect performance across all participants, regardless of learner type, we used the full sample (N = 145) in this analysis. Further, because points were awarded only when participants shipped units of office equipment, either on a truck or out through the pick-up window, two task actions essential to the successful completion of the task were “shipped truck” and “sent out pick-up.”

If participants learned to more-efficiently complete the task with task practice, we would expect to see an effect of task practice (first vs. final trial) on both the time participants took to dispatch an initial truck and on the time participants took to dispatch an initial pick-up order shipment. In support of Hypothesis 1, there was a significant effect of trial on initial truck shipment speed, $F(1, 144) = 261.45, p < .0001, \text{Wilks’ Lambda} = .35$, and on initial pick-up order shipment speed, $F(1, 144) = 153.90, p < .0001, \text{Wilks’ Lambda} = .48$. The means for participants’ speed of initial task actions in Trial 1 and Trial 5 are displayed in Figure 1. Participants performed initial task actions more rapidly in Trial 5.
than in Trial 1.

We also expected that increased speed of initial truck shipment and pick-up order shipment would distinguish between those participants demonstrating the greatest learning and those demonstrating the least learning (Hypothesis 2). To test this prediction, we conducted a 2 x 2 mixed factorial analysis, with one between-subject variable (learner type) and one within-subject variable (trial). We used the reduced sample (n = 97) in this analysis, which included 49 participants in the greatest-learning group and 48 participants in the least-learning group.

Contrary to expectations, those in the greatest-learning group did not display significantly different improvements in their speed of initial truck shipment from the first to the final trial, compared to those in the least-learning group; performance improved similarly with task practice for both groups. Learner type had a significant main effect on speed of initial truck shipment, $F(1, 95) = 8.26$, $p = .01$, Wilks’ Lambda = .38. There was also a significant trial effect, $F(1, 95) = 153.46$, $p < .0001$, Wilks’ Lambda = .38, in the reduced sample, consistent with the results for the full sample used in testing Hypothesis 1. However, the learner type by trial interaction effect on speed of initial truck shipment was not significant. The means for the greatest-learning and the least-learning groups’ speed of initial truck shipment in Trial 1 and after task practice in Trial 5 are displayed in Figure 2.

In contrast, those participants in the greatest-learning group improved significantly more in speed of initial pick-up order shipment from the first to the final trial, relative to those in the least-learning group. The learner by trial interaction effect on speed of initial pick-up order shipment was significant, $F(1, 95) = 7.70$, $p = .01$, Wilks’ Lambda = .92, providing partial support for Hypothesis 2. There was also a significant effect of learner type on speed of initial pick-up order shipment, $F(1, 95) = 29.54$, $p < .0001$, as well as for trial, $F(1, 95) = 100.90$, $p < .0001$, Wilks’ Lambda = .48. The means for the greatest-learning and the least-learning groups’ speed of initial pick-up order shipment in Trial 1 and Trial 5 are displayed in Figure 3.

**Information-seeking Behavior as an Indicator of Learning (Hypothesis 3)**

To examine the relationship between information-seeking behavior and learning, we examined the relationship between total number of rule call-ups during Trial 1 and final performance scores in Trial 5 using the full (N = 145) sample. Results did not indicate a significant correlation between rule call-ups during Trial 1 and Trial 5 performance scores, providing no support for Hypothesis 3.

**Errors as a Measure of Learning (Hypotheses 4, 5, and 6)**

To examine the relationship between initial errors and learning (Hypothesis 4), we examined the relationship between rule violations (errors) during Trial 1 and final performance scores in Trial 5 using the full (N = 145) sample. Results indicated significant positive correlations for Rule 2, $r = .29$, $p < .01$; Rule 5, $r = .21$, $p = .01$; and Rule 7, $r = .47$, $p < .01$, providing partial support for Hypothesis 4.

To test the prediction that types of errors would
distinguish between those individuals in the greatest-learning group and those in the least-learning group (Hypothesis 5), we compared first the individual rule violations made by the participants in the greatest-learning group to those of the least-learning group in Trial 1, and then the individual rule violations made by the greatest-learning group to those of the least-learning group in Trial 5. Because our analyses included the learner variable, we used the reduced (n = 97) sample. For Trial 1, results indicated that participants in the greatest-learning group made significantly more violations relative to those in the least-learning group for Rule 1, \( F(1, 95) = 5.15, p = .02 \); Rule 2, \( F(1, 95) = 6.90, p = .01 \); and Rule 7, \( F(1, 95) = 8.21, p = .01 \). The means for the greatest-learning and the least learning groups' specific rule violations during Trial 1 are displayed in Figure 4. For Trial 5, results indicated that the greatest-learning group made significantly fewer violations than the least-learning group for Rule 1, \( F(1, 95) = 36.90, p < .0001 \); Rule 2, \( F(1, 95) = 11.22, p < .01 \); and Rule 7, \( F(1, 95) = 77.29, p < .0001 \). However, during Trial 5, the greatest-learners also made significantly more violations of Rule 3, \( F(1, 95) = 8.69, p < .01 \) and Rule 5, \( F(1, 95) = 10.47, p < .01 \). The means for the greatest-learning and the least-learning groups' specific rule violations during Trial 5 are displayed in Figure 5. In summary, results indicated that certain rule violations (types of errors) distinguished between participants who demonstrate the greatest learning and those who demonstrate the least learning, providing partial support for Hypothesis 5.

To examine our prediction that the greatest-learning group and the least-learning group would demonstrate different patterns of change in errors (Hypothesis 6), we conducted a 2 x 2 mixed factorial analysis with one between-subjects factor (learner type) and one within-subjects factor (trial). We again used the reduced (N = 97) sample. We observed a significant trial by learner type interaction effect for Rule 1, \( F(1, 95) = 38.02, p < .0001, \text{Wilks' Lambda} = .71 \); Rule 2, \( F(1, 95) = 19.84, p < .0001, \text{Wilks' Lambda} = .83 \); Rule 3, \( F(1, 95) = 8.87, p < .01, \text{Wilks' Lambda} = .91 \); Rule 5, \( F(1, 95) = 9.59, p < .01, \text{Wilks' Lambda} = .91 \); and Rule 7, \( F(1, 95) = 66.82, p < .0001, \text{Wilks' Lambda} = .59 \), providing partial support for Hypothesis 6 in that changes in errors distinguished between participants demonstrating the greatest learning and those demonstrating the least learning.

**Discussion**

Our study had two main purposes. The first purpose was to determine if alternative measures of learning could be applied to training on our truck-dispatcher task. Our second purpose was to identify specific indicators of learning that occur during training, which could then be used to detect learning and to modify training parameters to improve training success. Using our entire data set, we evaluated the relationship between certain predictors (i.e., task speed, information seeking, and number of initial mistakes) and final performance. Using a reduced data set, we evaluated the effects of learner type on task speed and patterns of mistakes. Results indicated that both the speed with which participants took to complete an initial task action, such as shipping out an initial truck or dispatching an initial pick-up order, as well as the nature and pattern of mistakes, could be used as indicators of learning. Our results suggest that giving participants more specific information about these two dimensions of performance could have beneficial effects on training outcomes in similar tasks. Moreover, our results...
highlight the benefits of using multiple, alternative behavioral variables other than final performance scores as indicators of learning.

**Task Speed as an Indicator of Learning**

As expected, our results indicated that, overall, participants’ speed of task actions improved with task practice for both dispatching initial trucks and pick-up orders. However, in complex tasks, some individuals may rapidly increase performance speed whereas others improve more slowly, sometimes not progressing beyond even a novice level (Ackerman & Beier, 2007). In the truck-dispatcher task, elapsed time taken to dispatch an initial pick-up order appeared to be an effective indicator variable for learning because it distinguished between participants who displayed the greatest amount of learning and those who displayed the least amount of learning. More specifically, we observed a significant trial by learner type interaction effect on time taken to dispatch an initial pick-up order. Those in the greatest-learning group became much faster at this skill by the final trial, relative to those in the least-learning group. This interaction effect was not observed for time taken to dispatch an initial truck. Dispatching trucks is possibly less challenging, as might be indicated by the shorter overall action completion times, whereas dispatching pick-up orders is more difficult (see Figure 1). Thus, speeding up pick-up order shipments could be a secondary strategy that further enhances performance after truck shipment is mastered. Future research should be done to further examine the effects of strategies on task understanding.

**Information Seeking as an Indicator of Learning**

Contrary to our hypothesis, our correlational analyses did not reveal a relationship between the amount of time participants spent re-reading task rules and final task performance. One possible explanation for this is that participants who focused on reading task rules during Trial 1 rather than immediately exploring the task did not gain as much actual task experience as others. Researchers (e.g., van der Linden et al., 2001) have noted that one type of behavior pattern that can arise when people are faced with a new and complex task is becoming too focused on obtaining more information, rather than exploring the task itself. Although having some information about a task is necessary for performance, if the new information leads to more uncertainty about how to approach the task and/or more information seeking, decisions and actions may be delayed or not taken at all, and participants could lose precious time to practice the task and/or learn from mistakes. Future research should determine when and how information-seeking behavior becomes detrimental to performance.

**Errors as Indicators of Learning**

We did observe significant correlations between certain rule violations at the beginning of training (Trial 1) and performance scores at the end of training (Trial 5) using the full sample. This provided initial support for the hypothesis that trainees experiencing a greater number of errors early in training would gain more experience with the task and would demonstrate better final performance than those who did not experience these errors early on. These significant correlations led us to further examine participants’ individual errors. However, we note that caution should be used when interpreting the correlations from the full sample because significant correlations are more likely in larger sample sizes.

Our analysis of errors indicated that those participants who displayed the most learning did not necessarily make fewer errors relative to those who displayed the least learning after task practice. Instead, different patterns of errors and different patterns of change in errors distinguished between the greatest- and the least-learning groups. In order to understand these differences in change and patterns, a thorough analysis of the individual errors is required. When comparing the two groups, we observed two interesting patterns of mistakes. One pattern was the reduction of certain errors (Rules 1, 2, and 7). The other pattern was the apparent use of errors (Rules 3 and 5) as learning opportunities beneficial to performance. A summary of the various rules and their implications may assist in understanding our results.

For Rules 1, 2, and 7, those in the greatest-learning group originally made more violations relative to those in the least-learning group during Trial 1; then they improved on these specific rules, making fewer violations of Rules 1, 2, and 7 during Trial 5 relative to those in the least-learning group. Rule 1 indicated that participants must place units on a truck or send them to a pick-up order window based on the label associated with each unit. If participants placed an item meant for pick-up onto a truck or vice versa, the item would not ship out and the participant would lose points. Obeying this rule seems intuitively necessary for successful performance, especially because violating it does not allow delivery of units. The necessity to performance was demonstrated in that those individuals whose performance scores increased the most demonstrated a significant decrease in this error relative to those whose performance did not increase as much.

Rule 2 required participants to ship units labeled as “regular” within two task days (4 min in real time) and units labeled as “rush order” within one task day (2

---

**Notes:**

- Psi Chi Journal of Undergraduate Research | Summer 2010
- Copyright 2010 by Psi Chi, The International Honor Society in Psychology (Vol. 15, No. 2/ISSN 1089-4136)
min in real time). At the end of each working day, the number of “regular” and “rush orders” that remained in the queue rather than shipped out were counted and points were deducted. This rule is based on speed of performance. Those in the greatest learning group displayed a significant decrease in this error after task practice relative to those in the least learning group.

Rule 7 required participants to accept orders into their queues from the waiting list of orders within one task day (2 min in real time). Similar to Rule 2, this rule encouraged speed of task actions and distinguished between participants in the greatest-learning and the least-learning groups.

In contrast, those in the greatest-learning group actually violated Rules 3 and 5 more often than those in the least-learning group, indicating a potential strategic use of errors. Rule 3 required that participants dispatch trucks to no more than three out of five available shipment zones; and Rule 5 stated that participants were required to limit loads on each truck to that truck’s capacity. Whereas violating either of these rules resulted in points being subtracted from participants’ total score, ignoring these rules could enhance efficiency and total performance score. For example, if a participant dispatched a truck quickly to four instead of three zones or slightly overloaded a truck, he or she was able to deliver more units to customers, and his or her performance score benefited from this increase in productivity. Although the participant lost a few points for each rule violation, the lost points were more than compensated for by the rapid delivery of more fully-loaded trucks. These results are consistent with research examining performance quantity/quantity tradeoffs (see Austin & Bobko, 1985), research which has suggested that individuals might ignore or violate rules (i.e., make mistakes) if that allowed them to increase performance quantity.

Theoretical and Practical Implications
Learner type distinction. For the purposes of our analyses, we compared individuals who displayed the greatest increase in performance scores (top 33%) from early to final task practice to those individuals who displayed the least increase in performance scores (bottom 33%). This does not imply that those individuals who displayed the least increase had the lowest scores, in fact performance scores varied widely in both groups. We are also not ruling out the possibility that participants in the least-improvement group actually learned something during training as most participants demonstrated some improvement from beginning to end. However, we were interested in identifying patterns of behaviors indicative of a change in strategy use or task understanding, and those individuals who displayed the least improvement in performance score did not demonstrate extreme changes in their behavior. Our results highlight the feasibility of identifying task features that are likely to distinguish learning rates. Such information might be useful in designing and delivering training interventions. For example, in the truck-dispatcher task, knowledge of trainees’ specific errors or speed of pick-up order shipment might enable trainers to adapt the training environment to enhance learning.

Using errors strategically. Our analyses demonstrated that for improved task success, participants forfeited certain rules in favor of increased production. This is certainly an example of the classic speed/accuracy trade-off (Austin & Bobko, 1985). As participants’ performance scores increased, they became faster at shipping trucks and dispatching pick-up orders, and they violated time-based rules (Rules 2 and 7) less often. However, although they managed to improve their scores and dispatch more units to customers overall, quality of performance declined. Trucks were slightly overloaded (Rule 5) or were dispatched to more than three zones (Rule 3). Ignoring these rules in favor of increased quantity of units delivered may be a strategy that trainers could use to encourage increased performance speed when necessary. On the other hand, the ability to identify behaviors that indicate an individual is forfeiting quality for quantity may be beneficial for trainers who want to encourage both quality and quantity of performance for the long-term success of a company. Our results suggested that although task instructions did not emphasize quantity over quality or vice versa, participants made strategic decisions to sacrifice some quality to obtain gains in performance quantity. Trainers might be similarly able to strategically focus trainees on quantity, quality, or both, to obtain organizational goals.

Limitations
As with any study, there are limitations that should be addressed. First, the use of undergraduate college students as participants may make the ability to generalize findings to the intended population more difficult. College students may be less motivated to complete the experimental task than employees given training for their actual jobs. Due to time constraints, college students are also more likely to hold part-time rather than full-time jobs. However, our results should support generalizations to entry-level, college-educated employees in part-time positions. Additionally, our study included a much higher number of females than males, so generalizations based on our research to male-dominated professions should be made with caution.

Additionally, we were only able to obtain limited information about participants’ behaviors in the...
Learning Measures in a Truck-Dispatcher Task | Carter, Steele-Johnson, and Claflin

current study. First, the variables recorded in the task simulation limited our ability to identify specific indicators of learning. The task simulation software recorded information included in performance score calculation (e.g., time at which trucks or pick-up orders were shipped or rules violated as well as the number of these events) but did not include all intermediate steps (e.g., moving items onto or off of a truck or from one truck to another). Knowledge of these intermediate steps may have improved our ability to identify specific indicators of learning. Also, we were only able to use scores and information generated during the first and final training trials to which participants were exposed. Although this may have limited our ability to determine exactly when trainees were learning during the task, it did allow us to observe specific changes in behavior from initial to final task practice.

Another limitation, which also relates to future research possibilities, is our operational definition of learner type. There are many possible ways to measure learning, and for the purposes of this study, we compared those participants whose performance scores displayed the greatest increase to those whose performance scores displayed the least increase. This comparison of the “most-improved” participants to those who did not display such changes in performance highlights the need for researchers and practitioners to specifically define a final criterion when evaluating training outcomes. Should training be considered more effective if participants demonstrate high scores or much improvement in scores? Certain people are top performers from the beginning of training and display little change in performance, whereas others show dramatic improvement from start to finish. Researchers could potentially assert that a high final performance score reflects the greatest learning during training; alternatively, examining improvement over the course of training is a separate area of interest in training effectiveness evaluation (e.g., Mathieu, Tannenbaum & Salas, 1992). Indeed, Sackett and Mullen (1995) suggested that a focus on a minimum criterion level of performance might be more useful for trainers than the typical focus on a criterion based on change or improvement.

Conclusion

In conclusion, we tested the validity of using alternative behavioral variables other than final performance scores as indicators of learning. Specifically, results indicated that for the truck-dispatcher task, task action speed and patterns of errors could be used as learning-indicator variables. Our results contribute to the literature by providing evidence that researchers can identify multiple behavioral indicators of learning, based on both speed and accuracy, and these can be applied to training improvement and research. We also contributed to the literature by providing evidence that people sometimes use errors strategically. Finally, our study highlights the notion that researchers should carefully consider whether to base assessments of training on final performance scores or improvement in performance across training and that their decisions should reflect the goals of the research and/or training they are conducting. By using an interdisciplinary approach to identify alternative measures of learning, this study opens up new avenues for future researchers who could use similar strategies to identify indicators of learning for other tasks.

References


Carter, Steele-Johnson, and Claflin | Learning Measures in a Truck-Dispatcher Task


---

**Appendix**

**Task Rules**

**Rule 1.** Participants had to correctly identify whether units were to be placed on a truck, or sent to the pick-up order window based on the label associated with the unit.

**Rule 2.** Participants were required to ship units labeled as “regular” within two task-days (4 min in real time), and units labeled as “rush order” within one task-day (2 min in real time).

**Rule 3.** Participants were required to dispatch trucks to no more than three out of five available shipment zones.

**Rule 4.** Each of the units displayed were labeled with a particular business name. Units with the same business names were to be shipped together on one truck rather than dispersed between several trucks.

**Rule 5.** Each truck had a limited capacity, and participants were required to limit loads on each truck to that capacity.

**Rule 6.** Trucks exceeding their capacity by more than 20% were not dispatched.

**Rule 7.** Orders first appeared in the “queue,” which was a waiting list of orders not yet accepted. Participants were required to accept orders from the queue within one task day (2 min in real time).