INTERNATIONAL JOURNAL OF Training and Development

International Journal of Training and Development 22:4
ISSN 1360-3736
doi: 10.1111/ijtd.12139

Enhancing training with well-designed checklists

Keaton A Fletcher, Wendy L Bedwell, Sarah E Frick and Britany N Telford

The theory of mental workload suggests that job aids should be particularly useful if they provide resources for individuals without creating excessive, additional cognitive burden. We tested this proposition by examining the individual and interactive effects of task-based training and checklist design on training performance. Undergraduate students (N = 229) were randomly assigned to task-based or non-task-based training, and to one of three checklists or no checklist to aid training performance. The three checklists were (1) designed to provide low levels of detail but highly structured, (2) high levels of detail and highly structured or (3) high levels of detail with low structure. Results suggest checklists improve accuracy and also minimize psychological strain, yet at the cost of reduced speed. This suggests industries in which accuracy is critical to performance outcomes should consider how checklists, such as safety checklists, are designed. Implications for checklist design and provision are discussed.

Introduction

Between 30 and 40 per cent of employees report not receiving formal training (Frazis *et al*., 1998). Inadequate training is an oft-cited organizational stressor that has been linked to poor performance (e.g. Palo & Padhi, 2003) and psychological strain outcomes such as job dissatisfaction (e.g. Chiang *et al*., 2005). Organizations that invest in training often experience the 'transfer problem', wherein skills that were learned during training do not translate into measureable changes in on-the-job behavior (Blume *et al*., 2010). Thus, organizations have turned to cognitive aids (e.g. job aids, training aids) to boost transfer of training and improve learning and retention during and after training (Duncan, 1985). These aids have demonstrated mixed results (Taylor *et al*., 2005).

One of the most commonly used cognitive aids within training and performance is the checklist (Hales *et al*., 2008). Checklists are structured cognitive aids that present

❒ Keaton A Fletcher, Georgia Institute of Technology, USA. Email: fletcherk13@gmail.com. Wendy L Bedwell is at Pace Consulting Solutions, USA. Email: paceconsultingsolutionsllc@gmail.com. and Sarah E Frick, Email: sfrick@mail.usf.edu and Britany N Telford, Email: btelford@mail.usf.edu are at the University of South Florida, USA.

information required for successful execution of tasks (Hales *et al*., 2008). The main function is to reduce human error to ensure a desired level of task performance, such as decreased error frequency and increased speed (Degani & Wiener, 1993). For example, implementation of checklists in healthcare has led to decreases in wrong-site surgeries, unsafe utilization of equipment and infection rates, as well as increased collaboration, teamwork and debriefing efficacy (e.g. Borchard *et al*., 2012). There exists little certainty on how to best design checklists for use in applied settings; practitioners often receive mixed messages with minimal empirical support (e.g. Catchpole & Russ, 2015). A review on checklist implementation in healthcare points, in part, to varying quality of checklist design as a potential explanation for widely varying rates of checklist use and compliance (Borchard *et al*., 2012). Therefore, this study's objective is to examine the nature of the relationship between training and checklist use and design on performance and individual well-being to advance both theory and practice regarding these frequently used training aids. To achieve this goal, we first review previous literature on checklist and training design, usage and interaction; then, we explain our research design and present the results; and finally, we explore the implications for both theory and practice in our discussion, as well as the limitations of our study and avenues for future research.

Theoretical underpinning

One key area that has received much theoretical attention regarding checklist design is balancing comprehensiveness and usefulness (Weiser *et al*., 2010). Some argue that a highly detailed checklist can help users remember all necessary information by externalizing it to a cognitive aid; others argue high levels of detail may lead to bulky and essentially unusable tools; thus, the most effective checklist should be 'short and simple' (Winters *et al*., 2009, p. 8). One aim of this paper is to determine how to best manage this balance to maximize the efficacy of training.

We argue that the theory of mental workload can shed light on the primary factors of effective checklist use and the balance between detail and burden (e.g. Wickens & Tsang, 2015). Mental workload is the balance between overall cognitive demands from a task, or set of tasks, and overall cognitive resources available to the individual (Wickens & Tsang, 2015). Resources are the summation of the finite cognitive capabilities and external entities that reduce task requirements (e.g. Wickens & Tsang, 2015). These resources are seen as equally useful for all types of cognitive demands placed on the individual.

According to the theory of mental workload, as individuals encounter increasing demands requiring more or increased complexity of cognitive functioning, perceived mental workload increases. When demands exceed available cognitive resources, individuals experience overload (Wickens & Tsang, 2015). Both heightened workload and overload are associated with decreased performance (Wickens & Tsang, 2015) and increased strain (Young *et al*., 2015). Yet, the theory argues this only occurs when demands outweigh resources.

Checklists are thought to reduce performance errors associated with compromised cognitive functioning by providing task-relevant information (Hales *et al*., 2008). Checklists may act as a resource upon which individuals can offload some of the cognitive burden associated with task completion (Fletcher & Bedwell, 2017). As a resource, then, checklists should positively impact performance and the psychological strain associated with performing. Although they may aid in reducing the cognitive burden of recalling steps or procedures, cognitive aids, such as checklists, require attention and may take time away from actual task execution, much like an interruption (Jett & George, 2003). We hypothesize:

Hypothesis 1: Those who receive a checklist will (a) make fewer errors, (b) require more time to complete a task, yet (c) report lower levels of psychological strain than those who receive no checklist.

Checklists may vary along two design dimensions: the amount of detail and the amount of structure provided (Fletcher & Bedwell, 2017). A checklist's level of detail and structure should determine the amount of cognitive resources and demand associated with its use. As an external cognitive aid, a checklist provides resources in the form of task-relevant details that the user does not need to remember (Fletcher & Bedwell, 2017). As the level of detail within a checklist increases, the amount of necessary recall decreases, which enables the checklist to act as a cognitive resource rather than cognitive drain.

Yet, the use of a checklist places demands upon the user by requiring time away from task – time to search, read and comprehend the actual content that is intended to be a resource. This level of demand may be exacerbated or mitigated by checklist structure. Catchpole and Russ (2015) highlight a variety of suggested structural perspectives (e.g. order based on function-type, group based on temporal or geographic proximity) all of which rely upon grouping information into some type of structure to improve usability. This information-processing mechanism, called chunking, emphasizes grouping like-content together (e.g. Gobet *et al*., 2001), and implies that as a checklist becomes more structured, the level of demand associated with its use should decrease. Low levels of detail, however, may not have enough information to counteract the demands associated with checklist use. Although low detail checklists lend themselves to higher structure because the content itself is tantamount to headers that might be found in high detail, high-structured versions, their lack of information forces users to spend time recalling information. Recall demands high levels of cognitive resource usage (Craik & McDowd, 1987), and if that information has not been crystallized, the potential for recall errors is high.

High levels of detail without structure are also likely to be cognitively burdensome and time consuming, as users must filter through potentially irrelevant information to find specifics. High detail, coupled with high levels of structure based on chunking should then serve to reduce the cognitive burden of both information recall and time away from task, as the chunking serves as a contextual cue, shown to aid recovery from interruptions (Hodgetts & Jones, 2006). This dual benefit, particularly in a timed task performance scenario, should maximize the resource benefit intent of using the checklist. Thus, psychological strain should be lowest in users of these maximally designed checklists. We hypothesize:

Hypothesis 2a: Use of a low detail checklist will increase errors as compared to high detail.

Hypothesis 2b: Use of a low structure checklist will increase task completion time as compared to high structure.

Hypothesis 2c: Use of a high detail, high structure checklist will decrease psychological strain as compared to a low detail, high structure checklist or a high detail, low structure checklist.

Task-based training as a resource

In line with the theory of mental workload, training can also be considered a type of resource as it provides individuals with the necessary information for successful task completion and goal achievement (e.g. Demerouti *et al*., 2001). Task-based training (TBT) provides resources in the form of declarative and procedural knowledge regarding proper task execution (Kraiger *et al*., 1993). It aids in creating accurate mental models (i.e. cognitive representations of phenomena), helping individuals to efficiently organize information for quicker recall (e.g. Cooke & Fiore, 2010). By providing individuals with more cognitive resources and the ability to better utilize these resources, TBT should reduce mental workload, thus increasing accuracy and decreasing time required to complete the task as well as strain (Wickens & Tsang, 2015). Empirical evidence across multiple contexts demonstrates that TBT results in increased performance (e.g. Taylor *et al*., 2005). We hypothesize:

Hypothesis 3: The provision of TBT will increase performance in terms of (a) reduced errors and (b) reduced time required to complete the task and (c) decrease psychological strain.

Training and checklists

Given that both training and checklists provide resources to individuals, it is necessary to determine how these potential resources interact so as to best complement one with the other. Meta-analytic evidence suggests the effectiveness of cognitive aid provision with training has been mixed (Taylor *et al*., 2005). This could be due to varied cognitive aid design or usage. As a cognitive aid, a checklist should help supplement and reinforce trained principles (Hales *et al*., 2008). To maximize checklist utility, information should be included on a checklist if it is not entirely retained from training, the transfer environment is particularly demanding or the cost of forgetting a step is high (e.g. Hales *et al*., 2008).

Given the resource enhancing benefits of TBT (Taylor *et al*., 2005) checklists may be seen as superfluous or harmful by the user due to the aforementioned cognitive demands associated with use and seemingly redundant cognitive information. In such instances, checklists may be viewed as a demand rather than a resource, resulting in disuse, a phenomenon referred to as checklist fatigue (e.g. Hales *et al*., 2008). When training complex information/skills that must be implemented immediately (e.g. on-the-job training), trainees often have difficulty effectively transferring what was recently learned (Blume *et al*., 2010). The theory of mental workload suggests the effects of TBT may be enhanced by a checklist. And in cases where individuals are not able to be trained, checklist provision would be critical as it would be the *only* task-relevant resource available to the non-trained individual.

The relationships between checklist provision, training provision and outcomes of interest are nuanced. To encourage use, especially following TBT, a checklist should be clearly designed and present information that is not fully retained or not easily recalled (Hales *et al*., 2008). A poorly structured checklist that presents information in a cognitively burdensome manner should have limited benefits. This should be particularly pronounced for those who have received TBT and do not have as high of a perceived need for the checklist as others who may not have been trained. Checklists that favor structure over the provision of detail, however, may be particularly harmful to those who have not received TBT or received only non-task-based training (NTBT). As such, we hypothesize the following:

> *Hypothesis 4a: Participants who receive NTBT and either no checklist or the low detail checklist will make the most errors, whereas those who receive TBT and the high detail, high structure checklist will make the least errors.*

> *Hypothesis 4b: Participants who receive TBT and the high detail, low structure checklist will take the most time, whereas those who receive TBT and the high detail, high structure checklist will take the least time.*

> *Hypothesis 4c: Participants who receive NTBT and the low detail checklist, those who receive will TBT and the high detail checklist, or those who receive no checklist will report the highest levels of psychological strain, whereby those who receive TBT and the high detail, high structure checklist will report the lowest levels of psychological strain.*

Method

Participants

Undergraduate students ($N = 239$) enrolled in psychology courses at a southeastern university in the USA received course credit for participation. All participants were randomly assigned to one of eight, fully crossed conditions (training type: TBT or NTBT by checklist type: none, low detail, high detail, ideally designed). Six participants were removed from analyses because they did not complete the task in the allotted 35 min, and four were removed due to missing performance data (e.g. task completion time, number of errors), resulting in a final sample of 229 participants (76.9 per cent female, 47.8 per cent Caucasian, 14.5 per cent Black, 20.6 per cent Hispanic, 12.7 per cent Asian, 4.5 per cent other/did not report). Participants ranged in age from 18 to 47, (M = 19.92, SD = 3.45).

Materials

Sorting task

Participants were asked to sort cards according to a set of rules, much like sorting supplies in an emergency medical crash cart or stock room or sorting resumes based on specific criteria. Sixty cards from the card game *Blink©* were used for task completion. Each card had one of six symbol types (lightning bolts, triangles, rain drops, stars, flowers or moons) presented in one of six colors (blue, brown, green, gray, red or yellow). Cards also varied on the number of symbols, ranging from 1 to 5. Participants were instructed to sort the cards by symbol type, color and number, resulting in only two correct, equally likely, solutions. A correct solution required the cards to be sorted in a grid such that each row contained only one color. The rows of colors were aligned alphabetically by color name. Within each row, the cards were ordered such that the number of symbols on the card was in accordance with the following pattern: 1, 2, 3, 4, 5, 5, 4, 3, 2, 1. Additionally, the instructions required that several unique cards, such as the green two of stars card, be in a specific location on the grid and certain other specific cards be removed and set aside.

Checklists

There were four checklist conditions. Participants randomly assigned to the no-checklist condition did not receive a checklist at any point in the experiment. Information within checklists was ordered temporally, and when highly structured, was grouped under major temporal markers (e.g. the end of a type of task and beginning of the following sequence of tasks). Those in the low detail checklist condition received a checklist with low levels of detail that had organizing key points (e.g. *colors are in alphabetical order top to bottom*) – by necessity the low detail checklist is high structure. Participants randomly assigned to the high detail, low structure condition received a checklist with a high level of detail without any organizing key points (e.g. *BLUE in the TOP row. BROWN in the SECOND row. GREEN in the THIRD row*). The final checklist condition, high detail, high structure, consisted of a checklist with both key points (e.g. *colors are in alphabetical order top to bottom*) and explicit sub-points related to the key point (e.g. *BLUE in the TOP ROW, then Brown, then Green, then Gray, then Red, then YELLOW in the BOTTOM ROW*), thus combining high levels of detail with high levels of structure. See appendices for full checklists.

Measures

Effectiveness

The number of errors was calculated by counting the number of cards in the participant's solution that did not match the correct solutions. Thus, a high number of errors (maximum of 60) indicates poor performance.

Efficiency

Task completion time was measured in seconds using a stopwatch. Time began when the researcher said 'begin' and ended when the participant indicated task completion and had no further changes to the final card sort.

Psychological strain

Psychological strain was measured using the eight-item distress subscale of the Short State Stress Questionnaire (Helton, 2004). This subscale is a self-report inventory designed to measure the participant's current affective distress (e.g. 'I feel distressed' α = .87). Participants rated how strongly, in that particular moment, the item described their current state on a scale from *strongly disagree* (1) to *strongly agree* (2).

Table 1: Condition means and standard deviation

Training condition	Checklist condition	Errors M(SD)	Time to complete M(SD)	Psychological strain M(SD)
TBT	No checklist	40.79(*17.69*)	615.34(*344.80*)	2.24(*0.75*)
TBT	Low detail	16.72(*15.37*)	742.14(*335.49*)	1.67(*0.61*)
TBT	Ideal	6.27(*9.92*)	759.10(*259.47*)	1.66(*0.67*)
TBT	High detail	13.46(*13.30*)	958.68(*358.16*)	1.64(*0.55*)
NTBT	No checklist	48.88(*13.58*)	525.96(*218.60*)	1.92(*0.56*)
NTBT	Low detail	26.15(*14.82*)	976.37(*360.27*)	1.82(*0.60*)
NTBT	Ideal	14.45(*10.15*)	848.14(*242.10*)	1.79(*0.63*)
NTBT	High detail	18.52(*16.98*)	884.48(*300.68*)	1.61(*0.40*)

Table 2: Correlation matrix

	Variable	*N*	*M*	*SD*	1	2	3	4	5	6
1	Training[a]	229								
2	Checklist detail: No Checklist[b]	229			–.02					
3	Checklist detail: Low[c]	229			–.01	–.32*				
4	Checklist detail: Ideal[d]	229			–.00	–.33*	–.34*			
5	Time to complete	229	790.55	335.87	.07	–.37*	.11	.02		
6	Number of errors	229	22.72	19.48	.18*	.63*	–.04	–.38*	–.22*	
7	Psychological strain	228	1.79	0.62	–.02	.27*	–.04	–.07	–.09	.22*

$^*p < .05$; [a]1 = TBT, 2 = NTBT [b]0 = All other conditions, 1 = No Checklist; [c]0 = All other conditions, 1 = Low detail checklist; [d]0 = All other conditions, 1 = Ideally designed checklist.

Table 3: Results of multiple analyses of variance

Outcome	Source	SS	d.f.	MS	*F*	*p*
Effectiveness (errors)	**Training**	**3373.74**	**1**	**3373.74**	**16.63**	**<.001**
	Checklist	**38599.15**	**3**	**12866.38**	**63.41**	**<.001**
	Interaction	150.15	3	50.05	0.25	.86
	Within	44846.34	221	202.93		
	Total	204745.00	229			
Efficiency (time to complete)	Training	90995.28	1	90995.28	0.97	0.33
	Checklist	**3935715.97**	**3**	**1311905.32**	**13.93**	**<.001**
	Interaction	**984288.93**	**3**	**328096.31**	**3.48**	**.02**
	Within	20819057.26	221	94203.88		
	Total	168837035.00	229			
Psychological strain	Training	0.02	1	0.02	0.05	.83
	Checklist	**6.48**	**3**	**2.16**	**5.96**	**.001**
	Interaction	1.95	3	0.65	1.80	.15
	Within	79.77	220	0.36		
	Total	819.73	228			

Procedure

After giving informed consent, participants completed the demographics questionnaire. Participants were randomly assigned to either the TBT or NTBT condition. Participants in the TBT condition watched a 6-min training video that discussed and modeled the details of the sorting task and offered strategies for success. Neither TBT nor NTBT mentioned or targeted checklist use. Participants were also given 60 sec to practice sorting the card deck during the video. Participants who received NTBT watched a video of equal length that asked them to pay attention to their breathing and imagine what the cards would feel and look like. They were not offered strategies for task completion nor allowed interaction with the cards, but were given 60 sec to practice 'breathing normally'. This was to control for the effects of time spent thinking about the task.

After the training videos, all participants were given brief verbal instructions on how to complete the task. Participants in the three checklist conditions were then given their randomly assigned checklist to aid in their execution of the task. Participants in the no-checklist condition received no materials at this point. Those participants who were given a checklist were also given a marker to check off items on their assigned laminated checklist. All participants were informed that they (1) could not ask the experimenter any questions and (2) would be timed during their task. Immediately following task completion, participants completed the psychological strain measure.

Results

Condition means and standard deviations are presented in Table 1. Correlations are presented in Table 2. We provide effect sizes (Cohen's *d*) where appropriate (Cohen, 1988).

To test our hypotheses, we conducted three 2 × 4 analyses of variance (ANOVAs), one for each outcome of interest (Table 3). Results for number of errors, time required to complete the task and psychological strain are displayed in Figures 1, 2 and 3, respectively. To test hypotheses 1a-1c and 2a-2c, we examined the main effects of checklist on each outcome of interest. Checklist condition had a significant effect on the number of errors made; F (3,221) = 63.41, $p < .001$, $d = 1.86$. Results of a Tukey HSD post hoc test suggest that those who did not receive a checklist made significantly more errors than any other conditions (M =44.62, SD = 16.26), and that those who received the ideally designed checklist made significantly fewer errors (M = 10.29, SD = 10.77) than those who received the low detail checklist ($M = 21.27$, SD = 15.70) but not those that received the high detail checklist ($M = 16.12$, SD = 15.42), thus supporting hypothesis 1a and partially supporting hypothesis 2a. Results also supported a main effect of checklist condition on the time required to complete the task; F (3,221) = 13.93, $p < .001$, $d = 0.87$. Results of a Tukey HSD post hoc test suggest that those who did not receive a checklist took significantly less time to complete the task than any other conditions ($M = 573.09$, SD = 292.91), thus supporting hypothesis 1b. There were no significant differences in time required to complete the task between the remaining checklist conditions, thus failing to support hypothesis 2b. Results supported a main effect of checklist condition on psychological strain; F (3,220) = 5.96, $p = .001$, $d = 0.57$. Results of a Tukey HSD post hoc test suggest that those who did not receive a checklist reported significantly more psychological strain than any other conditions ($M = 2.09$, SD = 0.68), supporting hypothesis 1c. There were no significant differences in psychological strain between the remaining checklist conditions, failing to support hypothesis 2c.

To test hypotheses 3a-3c, we examined the main effects of training condition on each outcome of interest. Training condition had a significant effect on the number of errors made, thus supporting hypothesis 3a; F (1,221) = 16.63, $p < .001$, $d = 0.55$. Pairwise comparisons indicate that those who received TBT ($M = 19.25$, SD = 19.25) made significantly fewer errors than those that received NTBT ($M = 26.28$, SD = 19.15). Results failed to support a main effect of training condition on the time required to complete the task, thus failing to support hypothesis 3b; F (1,221) = 0.97, $p = 0.33$, $d = 0.13$. Lastly, results failed to support a main effect of training condition on psychological strain, thus failing to support hypothesis 3c; F (1,220) = 0.05, $p = .83$, $d = 0.00$.

To test hypotheses 4a-4c, we examined the interaction of checklist and training condition on each outcome of interest. Checklist and training condition did not interact to predict the number of errors made, failing to support hypothesis 4a; F (3,221) = 0.25, $p = .86$, $d = 0.11$. Results supported an interaction between checklist and training condition on the time required to complete the task; F (3,221) = 3.48, $p = .02$, $d = 0.43$. Post hoc Tukey tests reveal that participants who received NTBT and either a low detail checklist ($M = 976.37$, SD = 360.27) or high detail checklist ($M = 884.48$, SD = 300.68) or those who received TBT and the high detail checklist ($M = 958.68$, SD = 358.16) took significantly longer ($p < .05$) to complete the task than those that did not receive a checklist, regardless of whether they received TBT ($M = 615.34$, SD = 344.80) or NTBT ($M = 525.96$, SD = 218.60), thus partially supporting hypothesis 4b. Further, those who received NTBT and the ideally designed checklist took significantly longer to complete the task only compared to those that received NTBT and no checklist. Lastly, results failed to support a significant interaction between checklist and training condition on psychological strain, thus failing to support hypothesis 4c; F (3,220) = 1.80, $p = .15$, $d = 0.31$.

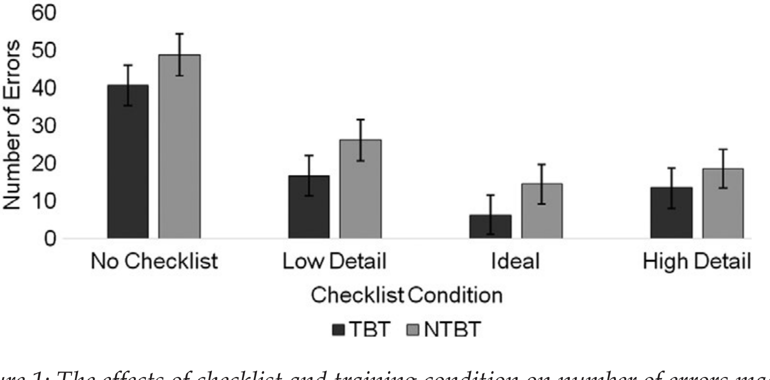

Figure 1: The effects of checklist and training condition on number of errors made. Note: Error bars represent two standard errors above and below the mean.

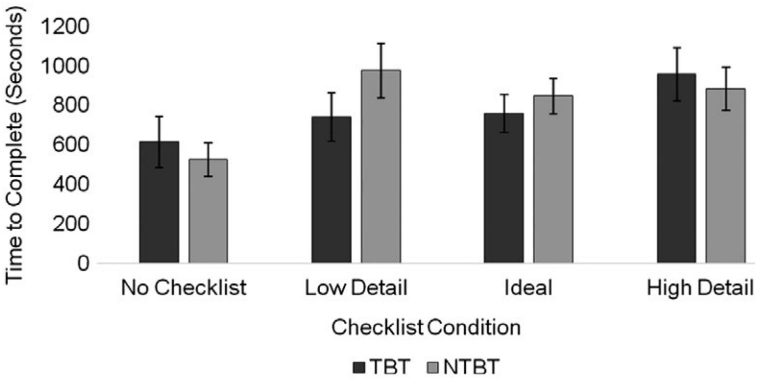

Figure 2: The effects of checklist and training condition on time required to complete the task. Note: Error bars represent two standard errors above and below the mean.

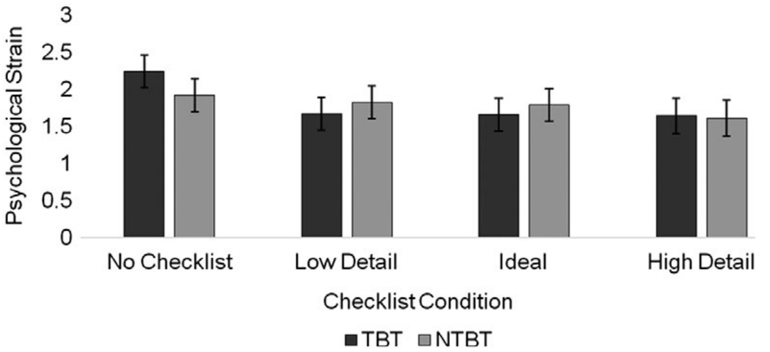

Figure 3: The effects of checklist and training condition on psychological strain. Note: error bars represent two standard errors above and below the mean.

Discussion

The purpose of this study was to determine whether checklists, as cognitive aids, interact with training to improve effectiveness and efficiency and decrease the psychological strain associated with performing. Informed by the theory of mental workload (e.g. Wickens & Tsang, 2015), we sought to examine whether the two key components of effective checklists, level of detail and level of structure (Fletcher & Bedwell, 2017), have an impact on performance. Results suggest use of checklists improves accuracy and minimizes the psychological strain associated with performance, but at the cost of reduced speed. Use of a checklist improves effectiveness, suggesting training performance errors can be minimized by checklist provision. This accuracy improvement comes at a cost of reduced speed, regardless of whether participants received TBT, particularly if the checklist is designed in a cognitively burdensome manner (e.g. high detail checklist without structure). Failure to provide a checklist to participants resulted in significantly higher levels of psychological strain. When considering whether to provide employees with a checklist to enhance training, the positive outcomes of improved accuracy and reduced psychological strain must be weighed against the cost of reduced speed.

When considering the pros and cons of checklist provision, particularly with regard to the speed accuracy trade-off, checklist design may offer a method of maximizing accuracy and speed. A theoretically ideal checklist (i.e. highly structured with sufficient detail) can provide necessary information but organize it in a way to minimize cognitive demand. Specifically, by organizing information, ideally designed checklists may improve accuracy, but not suffer the associated cost of reduced speed seen by similar checklists that provide necessary information but poor organization. In other words, an ideally designed checklist marries the resource provision of high detail checklists with the minimal cognitive burden associated with low detail checklists to maximize accuracy and minimize speed cost.

To better understand the cognitive underpinnings behind checklist design, we look towards the worst performing combinations, specifically NTBT and low detail, and TBT and high detail. Individuals who received NTBT made more errors and took longer to complete the task when they received the low detail checklist suggesting that it is the lack of necessary information that is problematic for this group. We also observe that participants who received TBT and the high detail checklist took significantly longer to complete the task, lending support toward the cost of cognitively burdensome checklist design. These individuals likely knew that they needed information from their checklist, and were eventually able to find it, but spent a significant amount of time searching for it. Taken together, this further lends support for the use of structure in checklist design, particularly in tasks where speed is critical.

It is worth noting, however, that participants did not report statistically significant differences in strain across different checklist designs. Many current evaluations of checklist design during development iterations (e.g. surveying users about the usability) ask users to evaluate usability or the ease-of-use (e.g. Degani & Wiener, 1993). These results suggest that perhaps participants are not as sensitive to the effects of design flaws as we might have expected.

Limitations of this study include the lab setting and student sample, which may not fully represent the complexities of employees in a work environment. Though this potentially minimizes the generalizability of the study, we believe that the findings should generalize to a wide variety of cognitively challenging tasks that are often undertrained. That said, performance contexts that may have greater personal significance or higher cost of error may show different results. Thus, we encourage future research into this finding, particularly with individuals in high-risk performance scenarios. Further, we only manipulated two dimensions of checklist design: detail and organizational structure. Other important variables, such as coloration or medium remain relatively unexplored. Future studies should examine whether these findings hold if the checklist is provided through a different medium (e.g. auditory or electronic) or if the tasks are less procedural and more fluid, like teamwork.

Despite these potential limitations our results suggest that checklists are promising tools that enhance accuracy following training, particularly if designed to balance cognitive demand with the provision of resources. Logically organizing key information increased the accuracy for both untrained and trained participants suggesting that checklists may be useful for both novices and more tenured employees. There are many high reliability industries, such as air traffic control, healthcare and chemical plants, in which accuracy is vital. Many of these industries already use checklists to ensure critical tasks are completed. Therefore, practitioners in these fields should examine existing and future checklist designs to ensure they are useful tools rather than cognitive burdens.

References

Blume, B. D., Ford, J. K., Baldwin, T. T. and Huang, J. L. (2010), 'Transfer of training: a meta-analytic review', *Journal of Management*, **36**, 4, 1065–105.

Borchard, A., Schwappach, D. L., Barbir, A. and Bezzola, P. (2012), 'A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery', *Annals of Surgery*, **256**, 6, 925–33.

Catchpole, K. and Russ, S. (2015), 'The problem with checklists,' *BMJ Quality and Safety*, **24**, 545–49.

Chiang, C. F., Back, K. J. and Canter, D. D. (2005), 'The impact of employee training on job satisfaction and intention to stay in the hotel industry,' *Journal of Human Resources in Hospitality and Tourism*, **4**, 2, 99–118.

Cohen, J. (1988), Statistical Power Analysis for the Behavioral Sciences, 2nd edn (Hillsdale, NJ: Erlbaum).
Cooke, N. J. and Fiore, S. M. (2010), 'Cognitive Science-Based Principles for the Design and Delivery of Training', in S. W. J.Kozlowski and E. Salas (eds), Learning, Training, and Development in Organizations (New York: Taylor & Francis), pp. 169–201.
Craik, F. I. and McDowd, J. M. (1987), 'Age differences in recall and recognition', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **13**, 3, 474–79.
Degani, A. and Wiener, E. L. (1993), 'Cockpit checklists: concepts, design, and use', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **35**, 2, 345–59.
Demerouti, E., Bakker, A. B., Nachreiner, F. and Schaufeli, W. B. (2001), 'The job demands-resources model of burnout', *Journal of Applied Psychology*, **86**, 3, 499–512.
Duncan, C. S. (1985), 'Job performance aids. Job aids really can work: a study of the military application of job aid technology', *Performance and Instruction Journal*, **24**, 4, 1–4.
Fletcher, K. A. and Bedwell, W. L. (2017), 'Cognitive Aids in Emergency Medical Services', in E. H. Lazzara and J.Keebler (eds), Human Factors in Emergency Medical Systems (Washington D.C: American Psychological Association), pp. 123–40.
Frazis, H., Gittleman, M., Horrigan, M. and Joyce, M. (1998), 'Results from the 1995 survey of employer-provided training', *Monthly Labor Review*, **121**, 3–13.
Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I. and Pine, J. M. (2001), 'Chunking mechanisms in human learning', *Trends in Cognitive Sciences*, **5**, 6, 236–43.
Hales, B. M., Terblanche, M., Fowler, R. and Sibbald, W. (2008), 'Development of medical checklists for improved quality of patient care', *International Journal for Quality in Health Care*, **20**, 1, 22–30.
Helton, W. S. (2004), 'Validation of a Short Stress State Questionnaire', in Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 48, No. 11). SAGE Publications, pp. 1238–42.
Hodgetts, H. M. and Jones, D. M. (2006), 'Contextual cues aid recovery from interruption: the role of associative activation', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **32**, 5, 1120–32.
Jett, Q. R. and George, J. M. (2003), 'Work interrupted: a closer look at the role of interruptions in organizational life', *Academy of Management Review*, **28**, 3, 494–507.
Kraiger, K., Ford, J.K. and Salas, E. (1993), 'Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation,' *Journal of Applied Psychology*, **78**, 2, 311–28.
Palo, S. and Padhi, N. (2003), 'Measuring effectiveness of TQM training: an Indian study,' *International Journal of Training and Development*, **7**, 3, 203–16.
Taylor, P. J., Russ-Eft, D. F. and Chan, D. W. (2005), 'A meta-analytic review of behavior modeling training,' *Journal of Applied Psychology*, **90**, 4, 692–709.
Weiser, T. G., Haynes, A. B., Dziekan, G., Berry, W. R., Lipsitz, S. R. and Gawande, A. A. (2010), 'Effect of a 19-item surgical safety checklist during urgent operations in a global patient population.' *Annals of Surgery*, **251**, 5, 976–80.
Wickens, C. and Tsang, P. S. (2015), Workload (Washington, D.C., US: American Psychological Association).
Winters, B. D., Gurses, A. P., Lehmann, H., Sexton, J. B., Rampersad, C. J. and Pronovost, P. J. (2009), 'Clinical review: checklists-translating evidence into practice,' *Critical Care*, **13**, 6, 210–19.
Young, M. S., Brookhuis, K. A., Wickens, C. D. and Hancock, P. A. (2015), 'State of science: mental workload in ergonomics,' *Ergonomics*, **58**, 1, 1–17.