

Review Article

A Systematic Review of Modifications and Validation Methods for the Extension of the Keystroke-Level Model

Shiroq Al-Megren , Joharah Khabti, and Hend S. Al-Khalifa 

King Saud University, Information Technology Department, Riyadh 12371, Saudi Arabia

Correspondence should be addressed to Shiroq Al-Megren; salmegren@ksu.edu.sa

Received 5 August 2018; Accepted 11 November 2018; Published 5 December 2018

Academic Editor: Thomas Mandl

Copyright © 2018 Shiroq Al-Megren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The keystroke-level model (KLM) is the simplest model of the goals, operators, methods, and selection rules (GOMS) family. The KLM computes formative quantitative predictions of task execution time. This paper provides a systematic literature review of KLM extensions across various applications and setups. The objective of this review is to address research questions concerning the development and validation of extensions. A total of 54 KLM extensions have been exhaustively reviewed. The results show that the original keystroke and mental act operators were continuously preserved or adapted and that the drawing operator was used the least. Excluding the original operators, almost 45 operators were collated from the primary studies. Only half of the studies validated their model's efficiency through experiments. The results also identify several research gaps, such as the shortage of KLM extensions for post-GUI/WIMP interfaces. Based on the results obtained in this work, this review finally provides guidelines for researchers and practitioners.

1. Introduction

Human-computer interaction (HCI) simplifies reality with models of human behaviour to design and evaluate computer systems [1]. Within HCI, models of motor behaviour lie on a continuum of analogy and mathematical equations. Generally, the models are categorised as either descriptive or predictive. Descriptive models present a framework to describe a phenomenon by identifying its features within a computer system. At the other end of the continuum, predictive models are commonly used to provide analytical a priori estimations of human performance without user participation, thus reducing time and resource consumption.

A family of predictive models (GOMS) were developed to compare and evaluate goals, methods, and selection rules of skilled, error-free user performances [2]. GOMS techniques model goal hierarchies of defined unit tasks rendered as a composition of action and cognitive operators [3, 4]. The simplest member of the GOMS family is the keystroke-level model (KLM), which predicts the execution time of specific tasks in a desktop environment using a mouse and keyboard. The KLM has been widely utilised to evaluate

expert performances of various desktop interfaces, and its aptitude and usefulness have been well demonstrated.

The challenges of designing and developing computer systems and the emergence of new technologies have revealed a need for updated quality assessments. Revising predictive models for these challenges can help evaluate human performance a priori and reduce the need for time- and resource-intensive human studies. The KLM was developed from and intended for desktop systems but has continually been extended to model systems designed for other computer setups in various domains. These extensions involve adapting the original KLM operators, introducing or inheriting new operators, revising heuristics, and presenting new execution calculations or techniques to satisfy the extension's purpose.

A systematic review of KLM extensions provides an objective procedure for identifying the extent of the research that is available; to the best of the authors' knowledge no prior systematic review exists that focuses on KLM extensions. This paper extensively reviews KLM extensions between 1980 and 2016. The goal of this review is to summarise, analyse, and assess the empirical evidence regarding the purpose for each extension, the extension's application domains and setups,

TABLE 1: KLM operators and predicted execution times in seconds [6].

Type	Operator	Description	Time (s)	
Physical	<i>K</i>	Keystroke or button press		
		Best typist (135 wpm)	0.80	
		Good typist (90 wpm)	0.12	
		Average skilled typist (55 wpm)	0.20	
		Average non-skilled typist (40 wpm)	0.28	
		Typing random letters	0.50	
		Typing complex codes	0.75	
		Worst typist	1.20	
		<i>P</i>	Point with a mouse to a target on a display	1.10
		<i>H</i>	Home hands on the keyboard or other device	0.40
Mental	<i>D</i>	Drawing n straight line segments with a total length of l	$0.90n_D + 0.16l_D$	
		<i>M</i>	Mental preparation for an action	1.35
System	$R(t)$	System response of t seconds	t	

wpm: words per minute.

and the research methods used to create and validate extensions. Most importantly, this review investigates how the KLM has been extended within the new models by examining operators, heuristics, equations, and domain-specific metrics. The results of this review also outline relevant issues for designers, developers, and researchers who apply or extend the KLM.

The rest of this paper is organised as follows. Section 2 presents the background for the KLM by introducing the topic and its seminal publications. Section 3 describes the methodology and protocol used to systematically review the KLM extensions. Section 5 describes the results of the review. Section 6 discusses the principal findings, limitations, and implications for research and practice. Finally, Section 7 concludes the paper and suggests future directions.

2. Keystroke-Level Model: An overview

KLM [6] is the simplest and most practical GOMS method for evaluating the time performance of user-computer system interaction. Underlying the KLM is the assumption that user employs a series of small and independent unit tasks. These tasks support the decomposition of larger tasks into manageable units. The sum of the durations of these small units equals the time it takes to complete the task. Each unit task has two phases: task acquisition and task execution: the total time to complete a unit task is the sum of these two parts:

$$T_{task} = T_{acquire} + T_{execute} \quad (1)$$

First, in the acquisition phase, the user conceptualises and develops a mental representation of the unit task. Then, during execution, the user invokes the appropriate system commands required to accomplish the unit task. The KLM predicts only the execution time of a unit task because that is the only phase over which a system designer has direct control.

Unit tasks in the KLM are described with a set of physical-motor, mental, and response operators (see Table 1). Operators are identified by a letter and include: *K* keystroking, *P*

pointing, *H* homing, *D* drawing, *M* mentally preparing, and *R* system response. *K* is the most frequently used operator and represents a keystroke or a button press. The operator *P* is the act of pointing to a target on a display with a mouse. *P* would typically be computed as a function of the distance to a target and its size (Fitts' law, [7]); however, for simplification it is assigned a constant time. In a typical computer setup, *H* is the action of moving the hand between keyboard and mouse and includes any fine hand adjustments on those devices. The physical operator, *D*, is restricted to the mouse and refers to manually drawing a set of straight-line segments within a constrained 0.56 cm grid. Before carrying out a physical action, the user has to mentally prepare for its execution. This preparation is represented by the *M* operator and a constant value of 1.35 seconds. The final operator, *R*, refers to the time it takes for the system to respond to a user's actions.

Unlike physical and system operators, *M* is not an observable user behaviour, yet it comprises a substantial fraction of the prediction. The occurrence of *M* is based on specific knowledge of user skills, and their placements are governed by a set of heuristics that embody psychological assumptions about users. Methods are a sequence of system commands that form a compiled segment of a user's behaviour when executing a unit task. A user cognitively organises a method according to cognitive chunks, and *M* typically occurs between chunks rather than within them. In Table 2, while Rule 0 identifies possible decision points within the methods, Rules 1 to 4 attempt to identify these method chunks.

Execution time is predicted by decomposing a unit task into a list of operators and then computing their summation:

$$T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R \quad (2)$$

$T_{operator}$ is an operator's total time, e.g., $T_K = n_K t_K$, where n_K is the number of keystrokes and t_K is the duration of each *K*. To illustrate how the KLM's equation and rules can be applied to predict user performance, consider the following example of a user renaming a folder to "klm" on a desktop. The user homes the hand on the mouse, *H*; points the mouse cursor at the object, *P*; double-clicks on the folder icon to

TABLE 2: Heuristic rules for M operator placement [6].

Rule 0	Insert M s in front of all K s that are not part of an argument string. Place M s in front of all P s that select commands.
Rule 1	When an operator following an M is fully anticipated in the operator just prior to M , remove the M .
Rule 2	When a string of MK s belongs to a cognitive unit, delete all M s except the first.
Rule 3	When a K is a redundant terminator, delete the M prior to the K .
Rule 4	When a K terminates a constant string, delete the M prior to the K ; however, do not delete the M when the K terminates a variable string.

allow for renaming, KK ; homes hands on keyboard, H ; keys new name “klm”, KKK ; and presses Enter, K

The KLM model without M and R (assuming an instantaneous response from the system) is $HPKKHKKKK$. Applying the heuristic rules for placing the M operators results in the final model $MHPKKHMKKKK$, where the first M is the time spent by user searching for the folder on the computer display, and the second M is the time the user requires to mentally prepare for typing. Therefore, (assuming K is 0.28 for average nonskilled typist):

$$T_{execute} = 2M + 2H + 1P + 6K = 5.8s \quad (3)$$

KLM was validated against observed values to determine how well the model predicted performance times and was subsequently used to model typical tasks in various systems (text editors, graphic editors, and executive subsystems). K 's value can be determined from a typing test prior to the test tasks. After a practice period, each expert user carried out test tasks and their keystroke times were logged. These times were then compared against the modelled predictions. The root-mean-square percentage error (RMSPE) was calculated as 21%. The developers of the KLM reported that this accuracy is the best that can be expected from the KLM and that it is comparable to the 20-30% previously obtained from more elaborate models [2, 6].

KLM inherits several limitations from GOMS. It assumes the user is an expert and does not account for user errors. This makes the model ill-suited for predicting average or novice system users. The model also assumes that the task is performed linearly; however, users often multi-task and are frequently interrupted. The KLM also does not consider individual differences in performance, such as mental workload and fatigue. In addition, the KLM predictive model is usually not generalizable because it is constructed to fit and evaluate a given interface.

3. Methodology

This systematic review of KLM extensions was carried out following the procedure given by Kitchenham and Charters [5, 8]. The review process consisted of three stages: planning, conducting, and reporting (see Figure 1). The review protocol was established after several meetings and discussions to reduce the risk of research bias. The rest of this section describes the research questions and the subsequent steps undertaken to conduct the review.

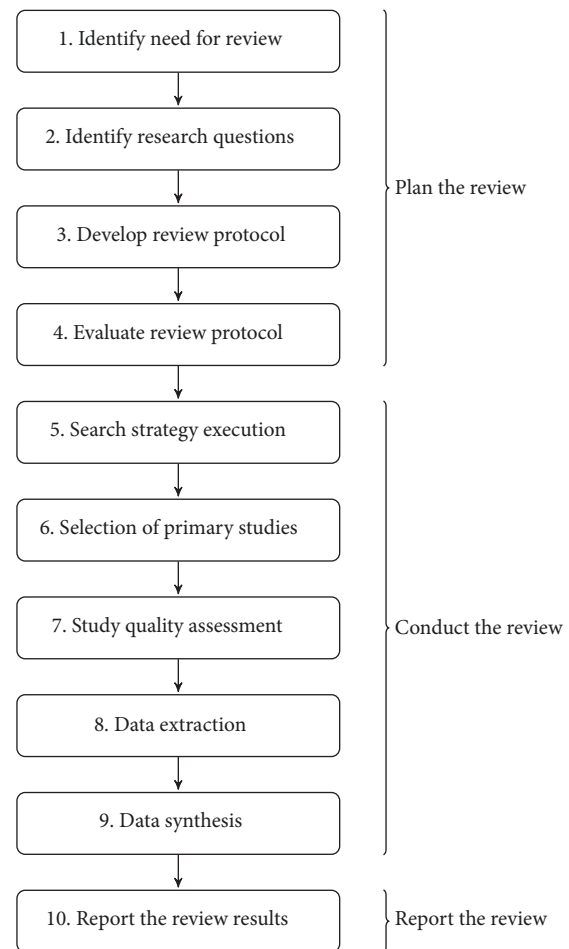


FIGURE 1: The systematic review process [5].

3.1. Research Questions. The goal of this review is to examine the current extensions of the KLM from the point of view of the following research question: “What extensions have been applied to the KLM and how have these extensions been developed and evaluated?” The question aims to summarise the current practices around extending the KLM to shed light on gaps in the current research, suggest areas for further investigation, and provide knowledge on the adoption of the KLM and its extensions to measure the performance of prototypes. Table 3 lists all the research questions and their motivations.

TABLE 3: Systematic review research questions and their motivations.

Research Questions	Motivation
RQ1: What was the purpose of extending the KLM?	Identify why the original KLM was modified, i.e., for which domain, application, or target user.
RQ2: What was the research method used to extend the KLM?	Identify research methods used to extend the original KLM, e.g., previous literature or user study.
RQ3: How was the original KLM extended?	Identify ways in which the original KLM was extended—specifically, how the operators and equations were affected. This includes, but is not limited to: original operators preserved, adapted or discarded, new operators and/or equations, updated M heuristics, and domain-specific metrics.
RQ4: What was the research method used to validate the extended KLM?	Identify research methods used to validate the extended KLM and the performance metrics utilised.

TABLE 4: Search string used to conduct the review search in the digital libraries. The wildcard * is used to signify any character to include variations for each term.

KLM	“keystroke level model” OR “keystroke-level model” OR “keystroke model” OR “KLM-GOMS” AND
Extension	exten* OR modif* OR enhanc* OR assess* OR evalu* OR revis* OR adapt* OR refin*

3.2. *Data Sources.* The main electronic database sources used to search for primary studies included the ACM digital library, IEEE Xplore, Springer Link, Elsevier Science Direct, Web of Science, Scopus, Taylor and Francis online, and Google Scholar.

3.3. *Search Strategy and Terms.* The search string consisted of two main parts: the KLM and its extensions (see Table 4). The first part relates to studies utilising the KLM for extension or evaluation, and the second part relates to extensions. The terms were extracted from textbooks and research papers on the KLM. The search string was formed by incorporating alternative terms and synonyms using the Boolean “OR” expression. The two main search terms were then combined using “AND”.

The search was conducted by applying the search string to collections of article meta-data. The string syntax was adapted for application to each digital library and its restrictions. This review was restricted to the period from July 1980 (the first time KLM was presented in “The Keystroke-level Model for User Performance Time with Interactive Systems,” [6]) to December 2016.

In addition to the primary search strategy, backward and forward searches were conducted. For each selected paper, the references were examined for a backward search, while the “cited by” links provided by some of the digital libraries were analysed for the forward search. Finally, publications citing the original KLM paper were also searched.

3.4. *Study Selection Criteria.* Each primary study was evaluated for relevance against inclusion and exclusion criteria. A study was selected when it satisfied one of the following inclusion criteria:

- (i) Studies explicitly extending the KLM.

- (ii) Studies reporting evaluations that employed the KLM or its extensions in post-WIMP interfaces or new application domains.

- (iii) Studies combining the KLM or its extensions with other models.

Studies were excluded from the review when they met one of the following exclusion criteria:

- (i) Studies presenting KLM-like extensions that did not extend the KLM or an extended version.

- (ii) Studies presenting extension recommendations.

- (iii) Studies presenting KLM testing processes that were focused on determining the effectiveness of the KLM for evaluation.

- (iv) Studies modifying the KLM to create composite operators.

- (v) Studies presenting duplicate reports of the same study and that did not present new material.

- (vi) Studies not written in English.

- (vii) Unpublished studies, excluding technical reports and theses.

3.5. *Quality Assessment.* The selected primary studies were assessed for relevance and strength using a three-point Likert-scale questionnaire consisting of the following subjective and objective closed-ended questions (see Appendix A):

- (i) The study presents a clearly stated purpose for extending the KLM.

- (ii) The study extends the KLM with new operators or modifications.

- (iii) The study clearly defines the research method used to extend the KLM.

TABLE 5: Digital search results resulting from applying the search string to the selected digital libraries. The numbers in parentheses show the number of unique articles.

Digital Source	Potential	Preliminary Selection	Final Selection
ACM digital library	44	23	18
Scopus	62	29	17
Springer Link	212	22	16
Elsevier Science Direct	184	17	5
Web of Science	31	10	10
Taylor and Francis on-line	67	12	4
IEEE Xplore digital library	93	7	5
Google Scholar	1751	118	62
Total	2444	238 (149)	137 (66)

- (iv) The extension methodology is adequate and repeatable.
- (v) The extension results and findings are clearly stated.
- (vi) The study clearly defines the research methods used to validate the extended KLM.
- (vii) The extension validation methodology is adequate and repeatable.
- (viii) The validation results and findings are clearly stated.
- (ix) The study presents a comparative analysis of the extended KLM against the original KLM.
- (x) The paper has been cited by other authors and/or contributes to the literature.

Each question is ranked 1 (yes), 0.5 (partly), and 0 (no). The final quality score is the sum of these values. The maximum score is 10 and the minimum score is 0. The quality of each primary study was ranked by two researchers. After thorough reviews, discussions were conducted to reach a final decision about the inclusion of each study in the review.

3.6. Data Extraction. The data extraction strategy was used to provide answers to the research questions in Table 3. An extraction form was developed to ensure that consistent extraction criteria were used (see Appendix A). The information extracted included:

- (1) Title, author, year, and type of publication.
- (2) RQ1: the purpose of the extension, device setup, application domain, and the intended users.
- (3) RQ2: the research method used to extend the KLM.
- (4) RQ3: how the KLM was extended, including adapted operator unit times, new operators or equations, updated M heuristics, and domain-specific metrics.
- (5) RQ4: the research method used to validate the viability of the KLM extension, and the performance metrics used for validation, including any comparison of the extension's performance against KLM.

3.7. Data Synthesis. The objective of this step was to accumulate and combine facts and formulate responses to the

research questions. The extracted information was grouped, summarised, and tabulated based on the six separate tables according to the elements identified in the data extraction process: general information, quality assessment, and RQ1–RQ4. Each primary study was assigned a code to identify the reviewed studies. Furthermore, extracted information concerning how the KLM was extended (RQ3) was collected in a table that lists and collates all the operators utilised.

4. Conducting the Review

Applying the review protocol yielded the preliminary results shown in Table 5. In total, 149 studies were selected. Using the defined inclusion and exclusion criteria, 62 primary studies (based on 66 articles) were identified. During this stage several issues were identified:

- (i) Some studies document different stages of the same research; for this reason, we refer to a smaller number of studies based on a larger number of articles.
- (ii) Some studies appeared in more than one source; these were considered based on the adopted search order (ACM, Scopus, Springer Link, Science Direct, Web of Science, Taylor and Francis on-line, IEEE Xplore, and Google Scholar).

The forward and backward search of the selected studies yielded only two relevant papers that were included. This low number indicates the thoroughness of the search terms used. The number of papers reviewed totalled 64 primary studies, based on 68 articles.

5. Systematic Review Results

This section summarises the results obtained after conducting the review and synthesis. First, an overview of the primary studies and their corresponding quality marks is presented. Next, the answers for three of research questions are addressed in separate subsections (RQ1, RQ2, and RQ4). Because research question RQ3 (see Table 3) is considered the most important, it is addressed in a separate section. Finally, a discussion and interpretation of the results is presented.

TABLE 6: Selected primary studies. The studies with a * were excluded based on the quality assessment results.

Study no.	Study	Study no.	Study
PS1	Roberts and Moran [9]	PS35	Dunlop and Montgomery Masters [10]
PS2	Kankaanpaa [11]*	PS36	Kulik, Kunert, Lux, and Fröhlich [12]*
PS3	Olson and Nilsen [13]	PS37	Sad and Poirier [14]
PS4	Edwards [15]	PS38	H. Li et al. [16]
PS5	Olson and Olson [17]	PS39	Liu and Rähkä [18]
PS6	Lane, Napier, Batsell, and Naman [19]	PS40	Overill, Silomon, and Chow [20]*
PS7	Quintana, Kamel, and McGeachy [21]	PS41	Schrepp [22]
PS8	Haunold and Kuhn [23]	PS42	Pettitt and Burnett [24]
PS9	Haunold and Kuhn [25]	PS43	Schneegaß, Pfleging, Kern, and Schmidt [26]
PS10	Koester and Levine [27]	PS44	Gokarn, Lobo, Gore, Doke, and Kimbahune [28]
PS11	Lee [29]	PS45	F. C. Y. Li, Guy, Yatani, and Truong [30]
PS12	Glenstrup, Engell-Nielsen, and Hansen [31]*	PS46	Holleis, Scherr, and Broll [32]
PS13	Manes [33]	PS47	Chatterjee, Sinha, Pal, and Basu [34]
PS14	Manes [35]	PS48	Webster [36]*
PS15	Green [37]	PS49	Spalteholz [38]
PS16	Dunlop and Crossan [39]	PS50	Jung and Jang [40]*
PS17	Bälter [41]	PS51	Rendon, Estrada-Solano, and Granville [42]*
PS18	Nowakowski [43]	PS52	Kang, Lin, Green, Pettinato, and Best [44]
PS19	Kieras [45]*	PS53	Rice and Lartigue [46]
PS20	Maragoudakis and Tselios [47]	PS54	Thompson III [48]*
PS21	Mori, Matsunobe, and Yamaoka [49]	PS55	El Batran and Dunlop [50]
PS22	Andrew Sears and Zha [51]	PS56	Lee, Song, Ryu, Kim, and Kwon [52]
PS23	John, Salvucci, Centgraf, and Prevas [53]*	PS57	Erazo, Pino, and Antunes [54]
PS24	Myung [55]	PS58	Palilonis [56]*
PS25	Jassar [57]*	PS59	Erazo and Pino [58]
PS26	How and Kan [59]	PS60	Jones, Wray, Zaiantz, Bachelor, and Newton [60]*
PS27	Jokela, Koivumaa, Pirkola, Salminen, and Kantola [61]	PS61	Green, Kang, and Lin [62]
PS28	Tonn-Eichstädt [63]	PS62	Elwart, Green, and Lin [64]
PS29	Hinckley et al. [65]	PS63	Burns and Ritter [66]
PS30	Holleis, Otto, Hussmann, and Schmidt [67]	PS64	Boring and Rasmussen [68]
PS31	Pettitt, Burnett, and Stevens [69]	PS65	Ginn [70]
PS32	Esteves, Komischke, Zapf, and Weiss [71]	PS66	Nasirinejad [72]
PS33	Hashizume, Kurosu, and Kaneko [73]*	PS67	Christensen, Pedersen, Bjerre, Pedersen, and Stuerzlinger [74]
PS34	Cox, Cairns, Walton, and Lee [75]	PS68	Erazo Moreta [76]

5.1. *Descriptive Statistics.* Table 6 shows the unique identifier assigned to each study and lists the associated reference. These identifiers will be used throughout the remainder of this review to refer to the primary studies.

5.1.1. *Quality Assessment.* Each quality assessment question was assigned a score of 1 (yes), 0.5 (partly), or 0 (no). The maximum score is 10 and the minimum score is zero. The quality scores were divided into categories:

- (i) Very High: $8 \leq \text{quality score} \leq 10$
- (ii) High: $5.5 \leq \text{quality score} \leq 7.5$
- (iii) Medium: $3 \leq \text{quality score} \leq 5$
- (iv) Low: $0 \leq \text{quality score} \leq 2.5$

Figure 2 demonstrates the quality-wise distribution of the primary studies. Of the 68 studies, 15 (22.06%) were

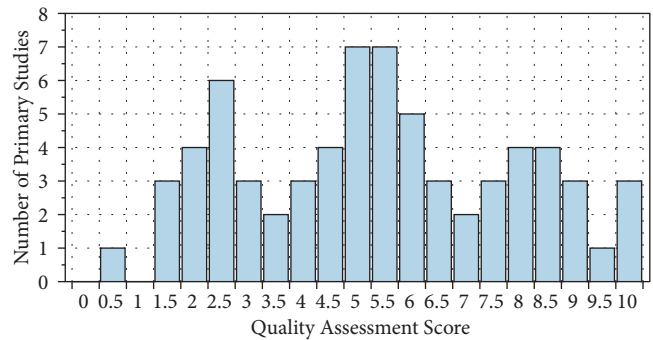


FIGURE 2: Quality-wise distribution of studies.

assessed as Very High, 20 (29.41%) were assessed as High, 19 (27.94%) were assessed as Medium, and 14 (20.59%) were

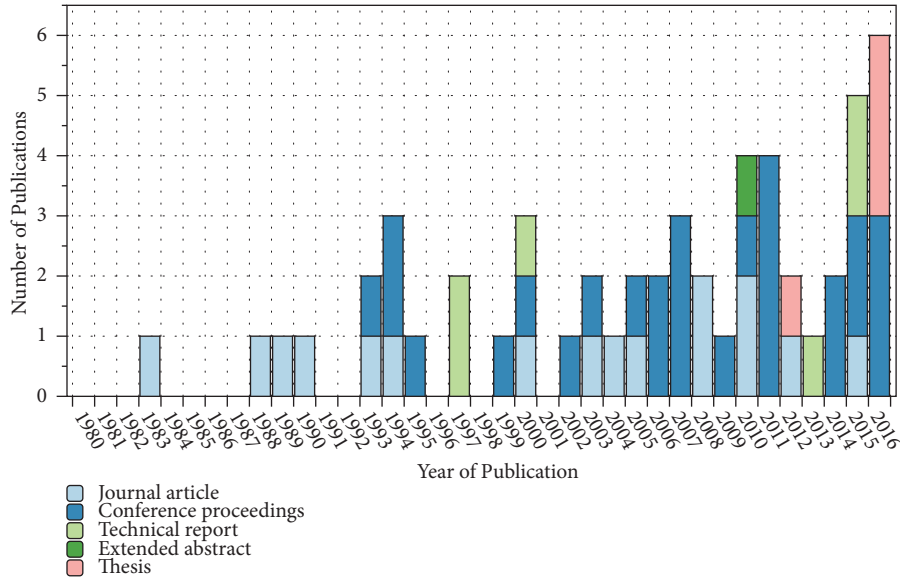


FIGURE 3: Year-wise distribution of studies and their publication type.

TABLE 7: Summary of top publications for KLM extensions.

Publication	Type	Number	Percentage
ACM Human Factors in Computing Systems	Conference	7	12.96%
University of Michigan Transportation Research Institute	Research centre	5	9.26%
HCI International	Conference	4	7.41%
HumanComputer Interaction	Journal	4	7.41%
Personal and Ubiquitous Computing	Journal	4	7.41%
IFIP Conference on Human-Computer Interaction	Conference	3	5.56%

assessed as Low. The primary studies PS3, PS14, and PS56 ranked at the top had quality scores of 10. Studies PS3 and PS56 were published in the journal of Human-Computer Interaction and Human Movement Science, respectively, and PS14 is a technical report from the University of Michigan Transportation Research Institute (UMTRI). The lowest score was 0.5 for PS48. The primary studies with low quality scores were excluded from analysis (PS2, PS12, PS19, PS23, PS25, PS33, PS36, PS40, PS48, PS50, PS51, PS54, PS58, and PS60), resulted in 54 primary studies with Medium, High, or Very High quality scores.

5.1.2. Publication Year. Figure 3 illustrates the distribution of published studies from July 1980 to December 2016. In the first decade after July 1980, only three (5.56%) journal papers were published in Communications of the ACM (PS1), Journal of Human-Computer Interaction (PS3), and the International Journal of Man-Machine Studies (PS4). The publication rate increased in the following decade (1990s), with 10 publications (18.52%, PS5-11, and PS13-15). This increase continued from 2000 to 2010, with 17 (31.48%) published studies (PS16-18, PS20-22, PS24, PS26-32, PS34-35, and PS37). A significant rise followed through the end of 2016 with 24 (44.44%) published studies (PS38-39, PS41-47, PS49, PS52-53, PS55-57, PS59, and PS61-68). This post-2010

spike in publications coincides with the resurgence of touch interactions and post-GUI configurations, which signalled a need for updated performance assessors.

5.1.3. Publication Sources. Table 7 summarises the details of the top publications for KLM extensions. Seven primary studies (12.96%) were published in ACM Human Factors in Computing Systems (CHI), this includes PS9, PS17, PS29-31, and PS38-39. Fewer than 10% of the primary studies were technical reports (PS13-14, PS18, PS52, and PS62) published through University of Michigan Transportation Research Institute (UMTRI). Just under 15% of the publications were journal articles: Human-Computer Interaction and Personal Ubiquitous Computing; the former had four primary studies published (PS16, PS27, and PS34-35). HCI International published four studies for 7.41% (PS29, PS32, PS37, and PS67). Three studies (5.56%) were sourced from the IFIP Conference on Human-Computer Interaction (PS11, PS46, and PS59).

Figure 4 illustrates seven of the eight digital sources used to search for primary studies and the number of publications retrieved. Overlaps existed between the various sources, and the studies were considered based on the search order. The results from Web of Science were encountered in other sources and are not shown. The majority (40.74%) of the

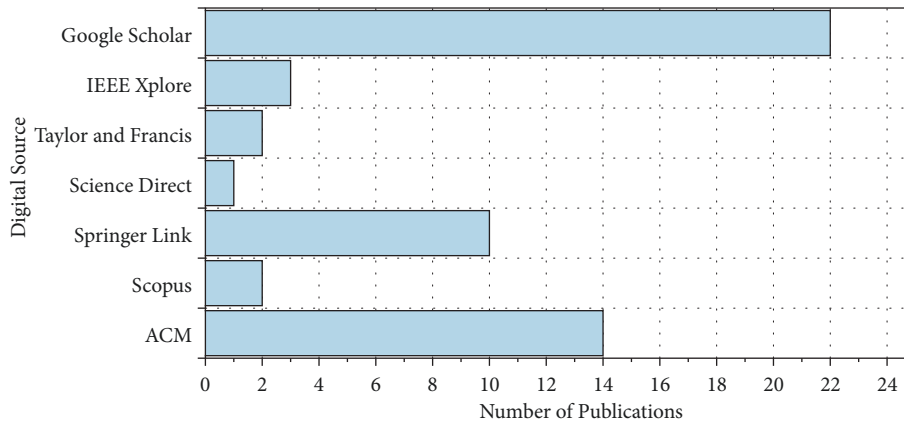


FIGURE 4: Distribution of primary studies across the seven digital sources utilised in this systematic review. The eighth digital source (Web of Science) is not shown due to overlapped content with earlier sources.

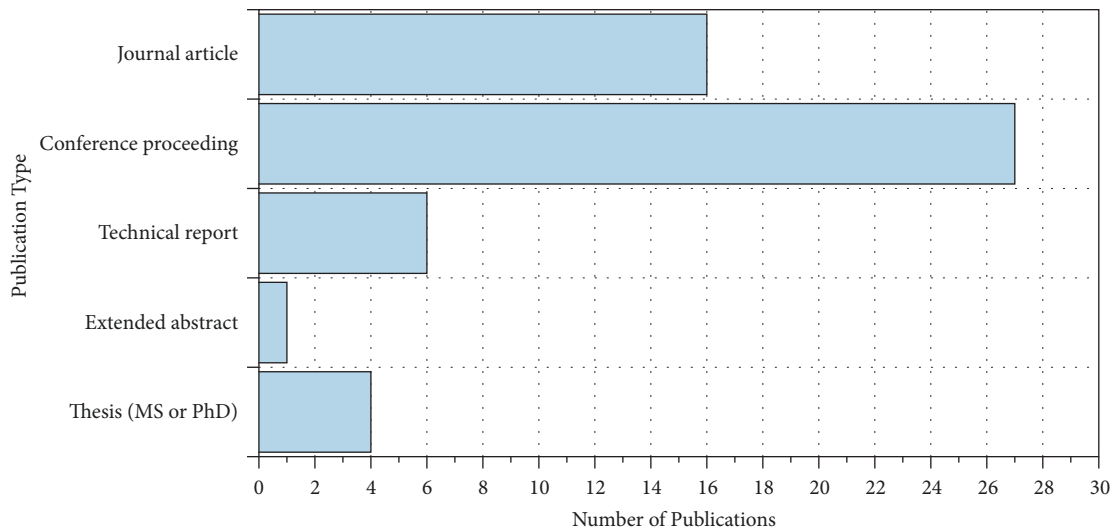


FIGURE 5: Distribution of primary studies across publication types.

studies were retrieved from Google Scholar. Fourteen studies (25.93%) were found in the ACM digital library. Springer Link produced 10 studies (18.52%), followed by IEEE Xplore with 3 studies (5.56%), Scopus and Taylor & Francis On-line with 2 studies (3.70%) each, and finally, a single publication from Science Direct (1.85%).

5.1.4. Publication Type. Publications were categorised as journal articles, conference proceedings, technical reports, extended abstracts, or theses. Figure 5 illustrates the distribution of primary studies across the five publication types. Statistics show that 50% (27 studies) of the studies were conference proceedings and 29.63% (16 studies) were in journals. The remainder of the studies were technical reports (6, 11.11%), theses (4, 7.41%), and one extended abstract (1.85%).

5.2. RQ1: What Was the Purpose of Extending the KLM? The motivation behind this research question was to identify why the KLM or one of its extensions was revised. This question

encompasses several areas of interests: main purpose, hardware technology (i.e., device setup), application domain, and intended users.

5.2.1. Main Purpose for Creating Extensions. Several purposes for creating extensions were identified in the review. These were collated into six main reasons:

- (P1) To fulfill a need for an updated quality assessment for new technologies, applications, techniques, or representations.
- (P2) To test the applicability of the KLM or its extensions.
- (P3) To extend the KLM to evaluate new technologies, applications, techniques, or representations.
- (P4) To revise the original or extended operators and heuristics.
- (P5) To extend the KLM or its extensions to describe additional interactions.
- (P6) To integrate KLM or its extensions with other models.

TABLE 8: Summary of the identified purposes of the selected primary studies and median quality scores.

Purpose ID	Number	Percentage	Median Quality Score
(P1)	20	37.04%	6.00
(P2)	4	7.41%	7.75
(P3)	16	18.52%	4.24
(P4)	12	18.52%	6.00
(P5)	7	9.26%	7.50
(P6)	8	7.41%	5.25

TABLE 9: Summary of the distribution of primary studies' device setups.

Device Setup	Number	Percentage
Traditional	13	24.07%
Mobile or tablet	20	37.04%
Key-based	12	22.22%
Touch-based	8	14.81%
In-vehicle Information System	10	18.52%
Traditional	6	11.11%
Touch-based	4	7.41%
Specialised	7	12.96%
Post-GUI	4	7.41%
Television	1	1.85%

Studies with (P1) as a purpose were conducted to extend the model, and in some instances validation studies were also conducted to confirm the viability of the model. For (P2), research methods were often utilised to extend and validate the extended model. In (P3), the studies typically extended the KLM or one of its enhancements and used experiments to determine the performance of a certain device or application. The fourth purpose, (P4), revised operators and heuristics using research methods; however, validation was infrequent. (P5) extended the model to describe new interactions, and utilised research methods to extend and validate the enhancement. The final purpose, (P6), utilised short studies to extend the KLM or one of its extensions to better incorporate it with a larger model.

Table 8 summarises the results of analysing the number of publications for each purpose. The table shows that publications with purposes (P2) and (P5) had the highest median quality scores (7.75 and 7.50, respectively). This is because these studies usually carried out experiments to extend and validate KLM. Publications that extended the KLM for purpose (P3) obtained medium quality scores (median of 4.24), since validation was often not considered.

5.2.2. Device Setup. Table 9 summarises the device setups collated from the review of 54 primary studies. The majority of extensions (20, 37.04%) modified the KLM or one of its extensions to model mobile or tablet interactions. These extensions were further categorised as key-based (12, 22.22%) or touch-based (8, 14.81%) mobile devices, smartphones, or

tablets. Fourteen studies (25.93%) extended the KLM for traditional configurations. The KLM was also extended for In-vehicle Information Systems (IVIS), which were categorised as either traditional (with knobs and dials, 11.11%) or touch-based (7.41.26%). Specialised configurations add features such as a digitized pad (PS8-9), Braille display (PS28), mouth-stick (PS10 and PS49), Leap Motion sensor (PS28), and specialised controls (PS64). Post-GUI configurations addressed extensions for natural user Interfaces (PS50, PS54, and PS68) and immersive projection (PS66). The KLM was also extended for web navigation on a television and for remote setup (PS47). Note that the percentages do not add up to 100% because one study (PS47) combined two setups.

5.2.3. Application Domain and Target Users. Several application domains were identified from the primary studies and grouped into high-level categories. Table 10 summarizes the recurrent domains. The most frequently examined domain relates to mobile or tablet applications (13 studies). Text and/or spreadsheet editing was the domain used to validate KLM [2, 6]; these studies were mainly conducted in the late 1980s to the early 1990s. Accessible interfaces were also examined to extend the KLM for interaction by blind users and users with motor disabilities. Navigating the web from various setups was also considered in the literature. IVIS setups were relatively popular (see Table 9) as a domain and considered tasks such as radio tuning, navigating lists, and using a global positioning system (GPS) for map navigation.

5.3. RQ2: What Was the Research Method Used to Extend the KLM? This research question examines the research methods used to modify the KLM or any of its extensions. The question is addressed in two ways:

- (1) What was the research method used to extend the KLM operators and heuristics?
- (2) What was the research method used to modify or compute the KLM operators' unit times?

Figure 6 demonstrates various research methods used to extend the KLM; these include experimentation, previous literature, observations, and several combinations of these methods. Twenty-seven (50%) of the studies did not use research methods to extend operators and modify heuristics. Observations were commonly conducted to identify or examine interactions (9 studies, 16.67%). Operators and heuristics were also extracted from previous literature (7 studies, 12.96%), and some studies used experiments (6 studies, 11.11%). Research methods were also combined. Several combinations were noted, including literature and experimentation, literature and observational studies, and observation and experimentation.

Figure 6 also illustrates the research methods used to modify the unit times of KLM operators. Of the 54 studies, only 12 (22.22%) did not utilise research methods. Over 50% of the primary studies (30 studies, 55.56%) conducted experiments to modify unit times. Eight studies (14.81%) relied on previous literature to adjust unit times. Additionally, research methods were also combined to extend unit times.

TABLE 10: Summary of recurrent domains from the primary studies.

Domain	Number	Studies
Accessibility	5	PS4, PS10, PS28, PS41, PS65
Games	1	PS56
Geographic Information Systems	2	PS8, PS9
In-vehicle Information Systems	10	PS13-15, PS18, PS31, PS42, PS43, PS52, PS6s, PS62
Medical	1	PS32
Mobile	13	PS16, PS20-22, PS24, PS26, PS30, PS35, PS35, PS37, PS39, PS45, PS46
Text and/or Spreadsheet Editing	4	PS1, PS3, PS4, PS6
Web navigation	4	PS11, PS17, PS47, PS49

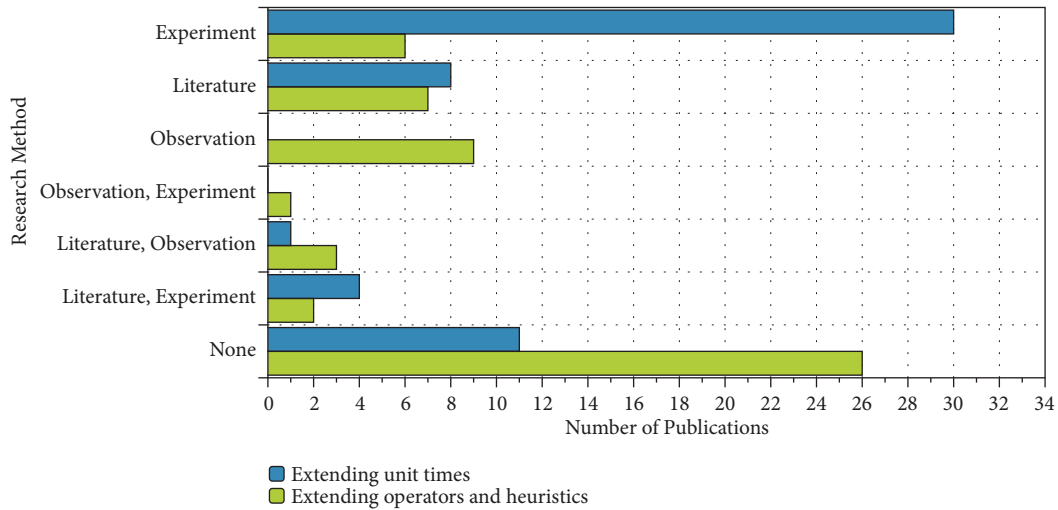


FIGURE 6: Research methods utilised to extend the KLM operators, heuristics, and unit times (RQ3).

These combinations include combining a literature search with either experimentation (4 studies, 7.41%) or observation (1 study, 1.85%).

5.4. RQ3: How Was the Original KLM Extended? The reviewed extensions demonstrated numerous ways in which KLM was revised:

- (i) Adapting the original KLM operators: K , P , H , D , M , and R .
- (ii) Inheriting operators from other KLM extensions or previous literature.
- (iii) Introducing new operators.
- (iv) Formulating new equations to calculate execution time.
- (v) Updating M heuristics.
- (vi) Computing domain-specific metrics.

It was also of interest to consider the operators that have been explicitly preserved in the extensions. The rest of this section discusses the operators, equations, heuristics, and metrics based on their intended device setup (see Section 5.2.2). Figure 7 collates the operators reported in the primary studies to identify their frequencies among the selected studies and device setups.

5.4.1. Traditional Setup. Thirteen studies were categorised as having a traditional setup: PS1, PS3-7, PS11, PS17, PS32, PS37, PS41, PS63, PS65. Figure 8 shows how the primary studies extended the KLM. The following sub-sections elaborate further on these operators, equations, and heuristics.

(1) Preserved Original Operators. K was the most popularly preserved of the original operators. PS6, PS32, PS11, PS65, PS41, and PS17 used the unit times associated with various typing skills. The majority of these used the time related to the speed of an average skilled typist (0.2 seconds), while others utilised the value 0.28 seconds (average non-skilled typist). H was preserved by PS1, PS11, PS32, and PS65, while M was used in three studies: PS1, PS17, and PS41. Four studies (PS1, PS32, PS41, and PS68) preserved the value of P . PS17 aimed to increase the accuracy of the P operator by utilising Fitts' Law. Operator R is system dependent and was often not utilised in the studies, yet it was still conserved.

(2) Adapted Original Operators. Some of the KLM operators were adapted through unit time adjustments or decomposition into finer tasks. PS7 updated the unit times of H , K , M , and P . PS63 dissected H into two actions: homing from the keyboard to the mouse and homing from the mouse to the keyboard. P 's unit time was updated in PS4, PS5, and PS7. A specialized P , $PM(l)$, was introduced in PS11 to indicate

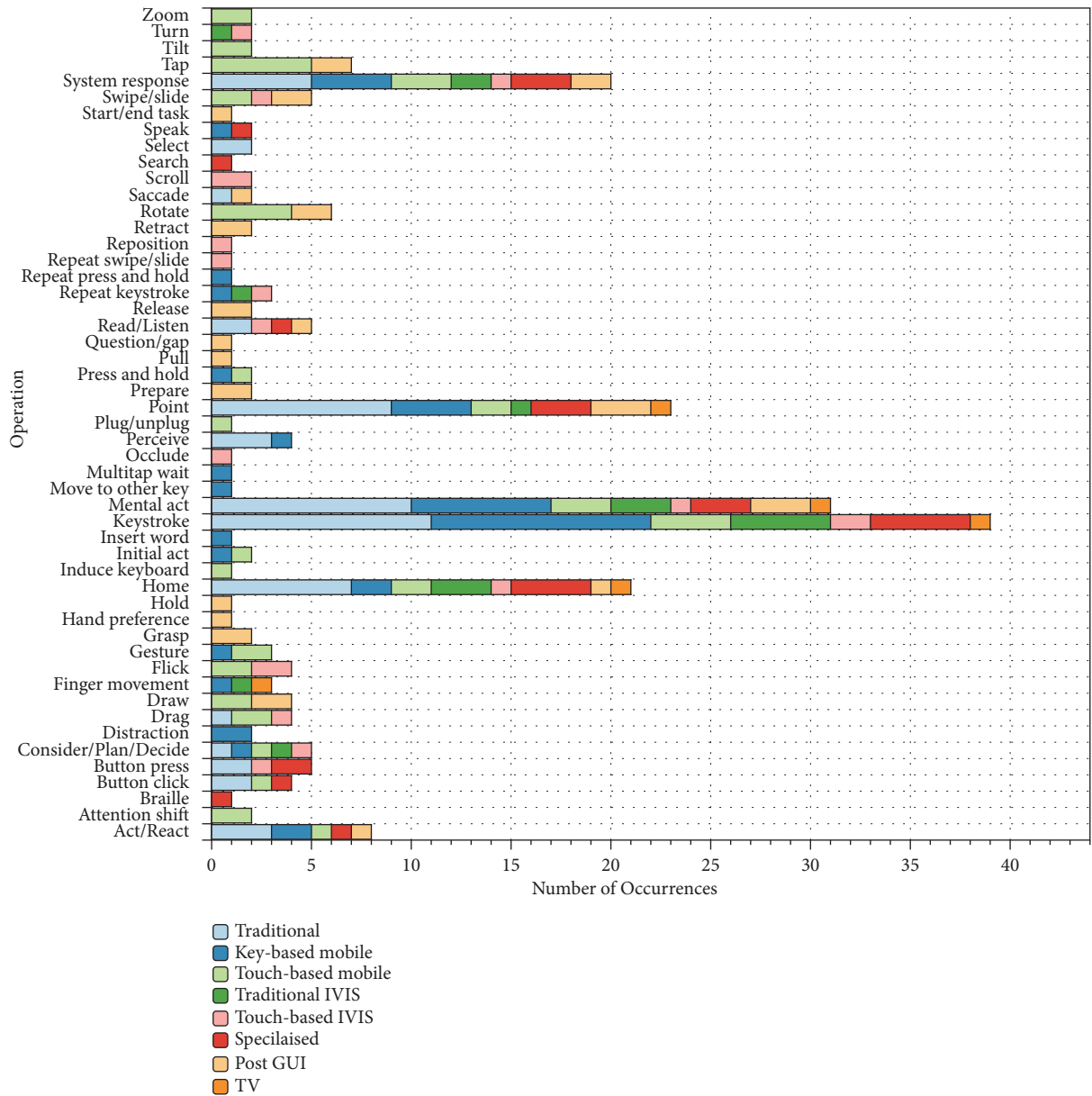


FIGURE 7: A bar chart of the collated operators and the frequency of adoption by the primary studies reviewed. The groups reflect the device setups previously discussed in Section 5.2.2: traditional, key-based mobile, touch-based mobile or tablet, traditional IVIS, touch-based IVIS, specialised, post-GUI, and TV.

pointing to the i th menu item. K and M were the two most frequently updated operators. K was updated in PS1, PS3, PS5, PS7, and PS63, while M was revised in PS3, PS5, PS6, PS7, and PS32. In PS5, M was decomposed into three mental actions: retrieval from memory, choosing among options, and executing a mental step. K was decomposed in PS3, where the unit times were 0.36 and 0.23 for two different spreadsheet tasks, respectively.

(3) *Inherited Operators.* PS11 inherited ten operators from previous extensions and prior literature: pressing a button B [45]; executing a mental step [17]; retrieving from memory, dragging to a menu item, and pointing to a menu item [77];

perceiving an image, reaction time of choosing an image, and eye movement [2]; menu search slope, intercept, and an overall value from an investigation into history tools for user support [52]; pressing a button and performing a button click [45].

(4) *New Operators.* New operators were introduced in 8 of the 13 traditional setup studies. PS1 identified several operators, including acquiring a task by looking at a certain manuscript A and using the arrow keys to point to a location Ps . PS7 introduced RW as the time required to read a word from the screen. PS4 identified several new operators: choosing a target, planning a route, moving to the next window, and

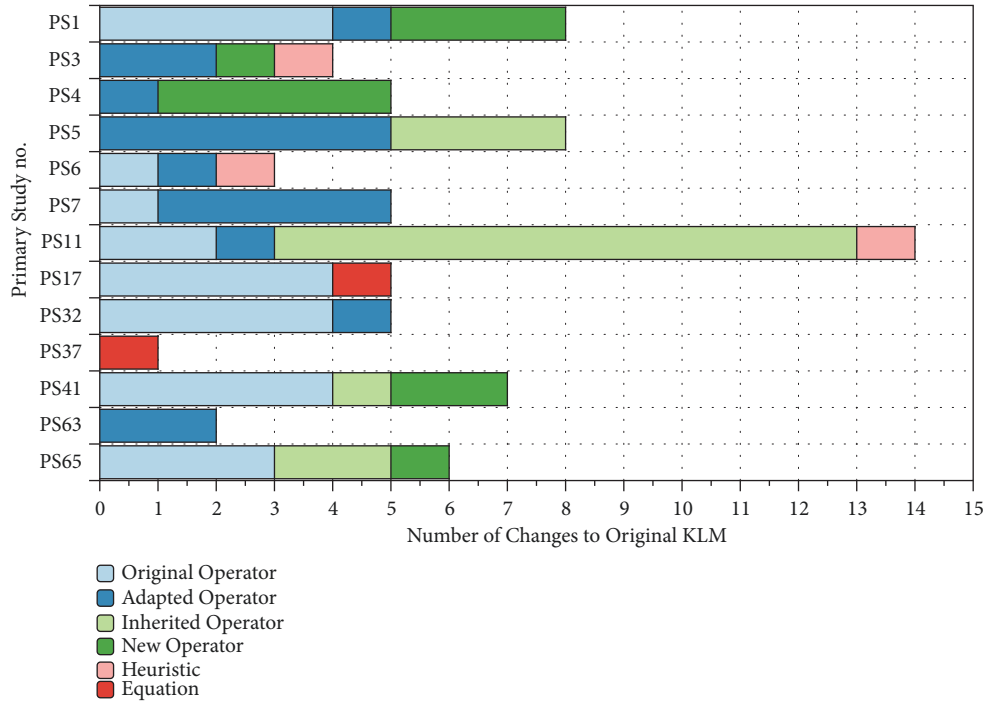


FIGURE 8: The number of original operators preserved, adapted, inherited from other extensions or prior literature, and newly introduced in the traditional setups category.

clicking the mouse. The symbol S was introduced in three studies (PS3, PS41, and PS51) to represent two different operations: mentally scanning/searching the display and pressing a keyboard shortcut. The time it takes to listen to a spoken word was utilised in PS64. PS41 also established a new operator for the time it takes to press a navigation key when navigating websites.

(5) *Updated Heuristics*. Heuristics are commonly updated when new operators are introduced to revise M placement. PS3 argues that commands issued through a series of menu choices involve a single M rather than one for each menu choice, because the command forms a single cognitive unit. Using a history tool, PS6 stated that switching from typing to using the history tool includes an additional long-term memory retrieval. For another history tool studied in PS11, M placement was extended for formula tasks. The study also offered guidance for placing the new mental scanning S operator.

(6) *New Equations*. In the KLM, task time is computed from the summation of the operators' unit times (see (2)). Some studies modified these equations to consider additional elements that may affect execution time. Both PS17 and PS37 introduced new equations in their extensions. PS17's authors formulated equations of various tasks that impact email archiving and retrieving. For word selection tasks, PS37 introduced equations to compute the time it takes to select a word given several variables, including scanning and scrolling time, word length, and the index of the selected word.

5.4.2. *Key-Based Mobile*. It was in the new millennium that interest in extending the KLM for mobile interaction and text entry became most evident. Twelve of the 54 selected studies modified the KLM to accommodate key-based mobile interactions (PS16, PS20, PS21, PS24, PS26, PS27, PS30, PS34, PS35, PS39, PS44, and PS46). Figure 9 illustrates these studies and the approaches they proposed to extend the KLM. The following subsections describe the changes made to extend the KLM.

(1) *Preserved Original Operators*. Several operators were preserved from the KLM: H , K , M , and P . PS16 utilised the original H , K , and M for predictive text entry on mobile phones. The KLM was further extended by the same authors in PS30 for five predictive text entry methods that preserved the original M and R . R was also used as is by PS21 and PS46. An extended KLM for modelling speech navigation and text entry preserved K .

(2) *Adapted Original Operators*. PS26 extended the KLM for SMS input by dissecting K into nine operators for various keys and repetitions. K was also decomposed in PS39 to reflect unique interactions with a Pinyin keyboard, an input method for Chinese text using the Pinyin method of romanisation. K 's unit time was revised in PS21, PS27, PS30, PS35, and PS44, the majority of which dissected the unit times based on the type of key and repetition. It should be noted that PS27 approached the KLM differently, assigning each key or repetition a score rather than a unit time. P has also been adapted and at times redefined. For instance, PS30 and PS46 modified P to reflect pointing with a device

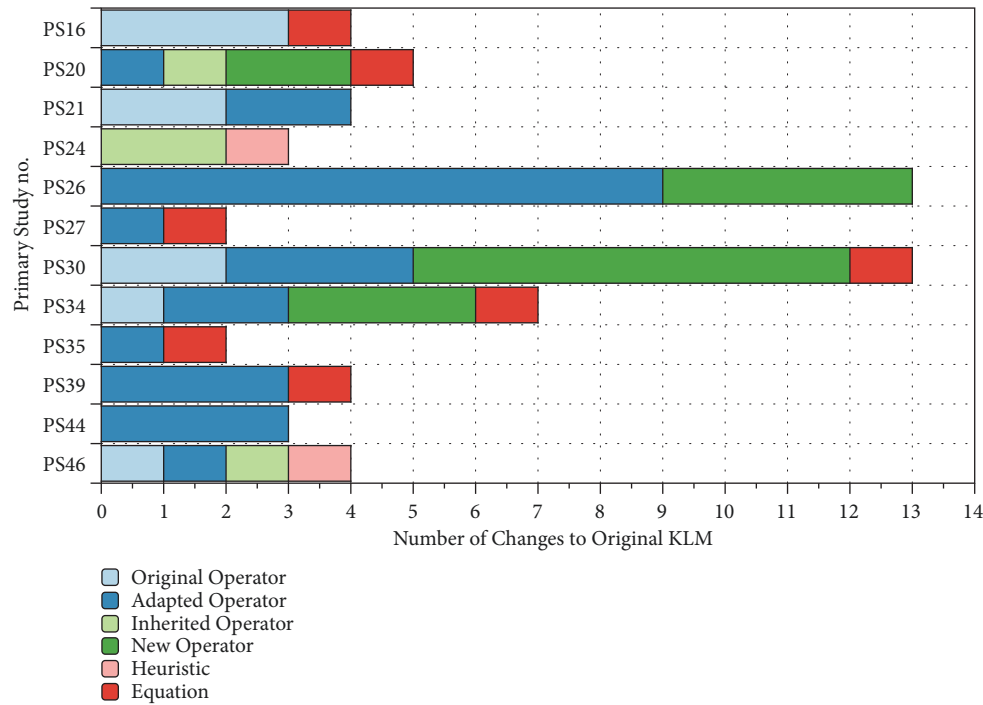


FIGURE 9: The numbers of original operators preserved, adapted, inherited from other extensions or prior literature, and newly introduced in the key-based mobile category.

to perform an action. PS44 considered P for pointing to a keypad. While PS21 preserved P 's original meaning, M was adapted in both PS21 and PS44 and decomposed in PS34 to represent time delays during text entry and recognition. TPER from PS20 is also an adaptation of M for text entry perception. H was revised for PS30 to consider the time needed to switch between listening/speaking on the phone and reading from the screen.

(3) *Inherited Operators*. Only four operators from three studies were inherited from the literature or another extension. Two of these replaced the value of K , the third updated the unit time for M , and the last re-used a value from a previous model for complex actions. PS24 enhanced the KLM to evaluate Korean text entry on a mobile phone where the values for K and M were inherited from Kim, Kim, and Myung [78] and John and Newell [79]. K 's unit time was also inherited from Silfverberg, MacKenzie, and Korhonen [80] to extend the KLM for message-text entry with a Greek corpus. Mobile KLM (PS30) was revised in PS46, which inherited the complex action operator to reflect tag-reading interactions.

(4) *New Operators*. Several new actions were recognised by half of the studies that extended the KLM for key-based mobile phones. PS20 introduced two new operators: waiting for the cursor to process when successive letters are entered from the same key in multi-tap text entry and the action of moving to another key. Similarly, PS26 utilised a wait operator for multitap entry. It also introduced MPHAlphaK (press and hold key), RPHAlphaK (repeat press and hold key), and InsertWord (insert word into corpus dictionary). Mobile

KLM (PS30) extended the KLM with several operators: attention shift for various focus shifts, complex actions, gesturing with phone, finger movement, initial act, and a multiplicative factor for distraction. PS34 extended the KLM for speech text entry and introduced an action that reflected the time needed to consider/recognise a command and utter a syllable.

(5) *Updated Heuristics*. A number of studies updated the placement of M and other perceptive actions to reflect interactions with a mobile phone. For Korean text entry (PS24), an M is expected to occur both before and after entering a syllable. Moreover, an M should not be placed before the next key since finger movement and the mental activity overlap. PS46 declared that M should appear before cognitive chunks and that an M is unnecessary before pointing at longer distances with respect to shorter ones.

(6) *New Equations*. When extending the KLM for text entry, new equations were commonly formulated based on the text entry techniques. Of the 12 studies, six modelled text entry on mobile phones (PS16, PS20, PS24, PS34, PS35, and PS39). PS16 introduced three new equations for traditional (multitap), predictive, and word-completion text entry with an English corpus. In PS35, the equation for predictive text entry from PS16 was reused for various word look-up techniques. For Greek text entry, PS20 formulated two equations to compare typical phone text entries against a newly developed approach. PS39 established an equation to compare the performances of two types of Chinese Pinyin input by integrating the KLM with other models. PS34 evaluated

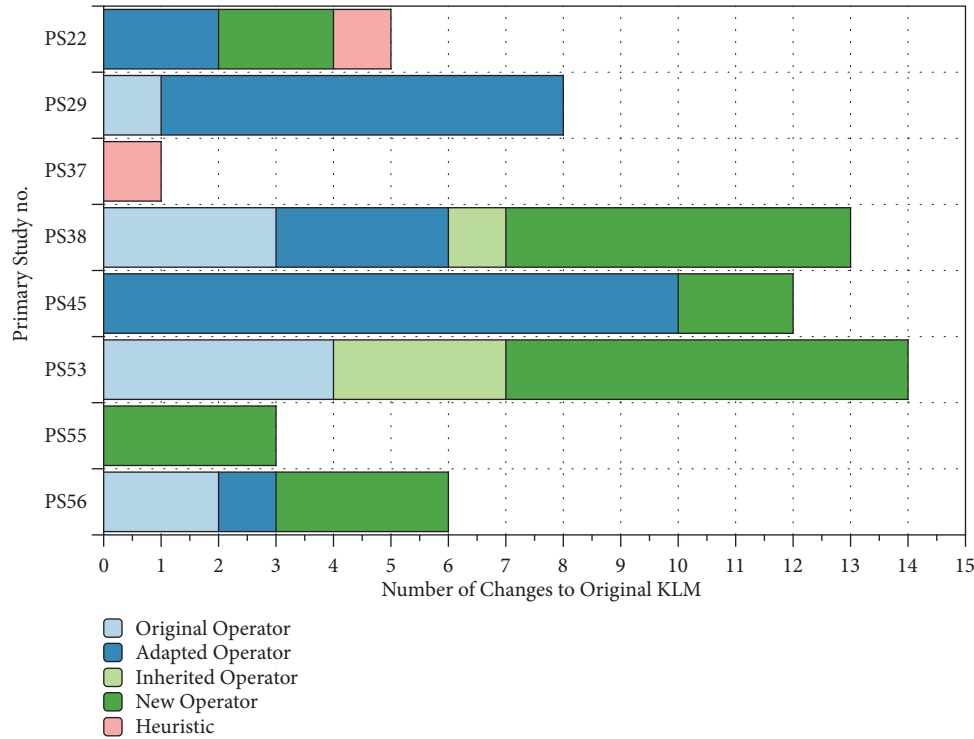


FIGURE 10: The number of original operators preserved, adapted, inherited from other extensions or prior literature, and newly introduced in the touch-based mobile or tablet category.

speech text entry compared with multitap and predictive text entry, in which several equations were constructed to consider time-out delays, number of words, and word options in predictive entry.

Two other studies formed equations in contexts other than text entry. The mobile KLM in PS30 proposed a new equation that took distractions of various severities into account. PS27 approached the KLM differently, presenting unit times as scores used to calculate the relative average efficacy, where the sum of the scores for each task is first divided by the number of tasks and finally multiplied by 100 to obtain a percentage.

5.4.3. Touch-Based Mobile or Tablet. Touchscreen interactions were considered in several KLM extensions as early as 2003. Eight studies from the selected 54 were identified (PS22, PS29, PS37, PS38, PS45, PS53, PS55, and PS56). Figure 10 illustrates how the new models were modified for extension. The following sub-sections describe the various updates applied to the KLM to represent touch-based mobile or tablet interactions.

(1) *Preserved Original Operators.* PS29 extended the KLM to measure the performance of a new interaction technique on a touch tablet under various conditions and with various styles. The model preserved R , as did PS53 and PS56. PS53 revised the KLM to accommodate a modern touchscreen interface, but preserved H , K , and M . The original operators D , H , and K were utilised by PS38 as an extended KLM for touch phone

mobile interactions. The primary studies PS53 and PS56 both preserved M , and PS53 also utilised the H and K operators.

(2) *Adapted Original Operators.* One KLM extension, developed to model the performance of a new interaction technique, decomposed P , D , and R . P was subdivided as follows: point stylus at segment, point to command, and point to end the mark. D_c and D_m symbolise drawing a circle around a dot and drawing a mark, respectively. R was divided into switching modes and the time it takes the system to respond. K was adapted by PS22 to consider both key repetition and movement between keys. In testing a new keyboard design for Chinese text input, 1 Line (PS45), K was dissected into a key for each finger on both hands. Similarly, M was modified in PS38 to reflect mentally initiating a task, deciding or choosing, retrieving, finding, and verifying. The extended model also adapts H into two actions: homing either a stylus or a finger to some location. PS56 modified P to reflect a relatively long movement from one position to another on a touch mobile phone in network gaming.

(3) *Inherited Operators.* PS53 inherited two operators from mobile KLM (PS30): initial act and distraction. Gesture actions were inherited but adapted to reflect the time needed to physically form specialised gestures with one or more fingers. The same operator was also used by PS38 to represent holding a gesture for a certain application.

(4) *New Operators.* Numerous operators have been created to represent touch interaction. PS53 extended the KLM

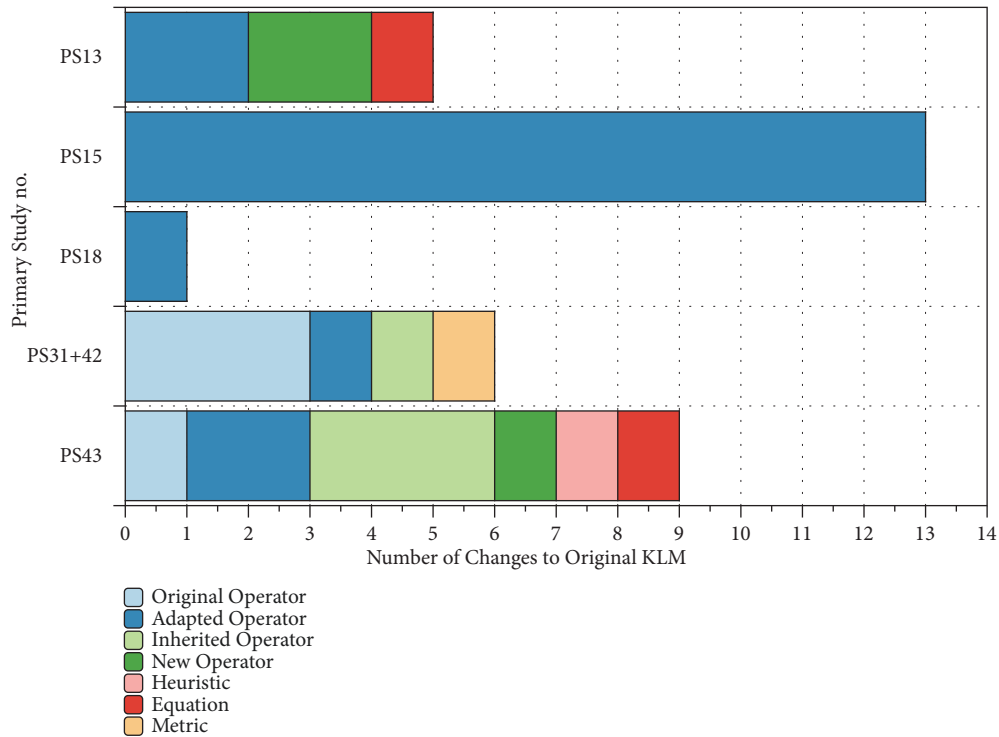


FIGURE 11: The number of original operators preserved, adapted, inherited from other extension or literature, and newly introduced in the traditional IVIS setup category.

to form a touch-level model (TLM) for touchscreen and mobile devices and introduced several new operators: tap, pinch/zoom to zoom in/out, swipe, rotate, drag element, and tilt device. Tapping is a common interaction in touch interfaces that was also introduced in PS38, PS55, and PS56. Swipe, zoom, and drag actions were also identified in PS55 and PS56. PS22 utilised two new operators that consider the decision and recovery times for data entry using a soft keyboard. Flick was established in PS56 to identify quick, short dragging actions. This action was decomposed in PS45 to distinguish between flick down and flick up. New operators introduced for finger/stylus touch mobile devices extended the model to include flipping or sliding a keyboard, continuously holding a key down, pressing a key on the side of the device, and plugging and unplugging other devices.

(5) *New Equations*. A quarter of the studies formulated new equations to compute the execution times of various tasks. PS22 formed a new equation for text entry on a soft keyboard that considers the number of characters, shifted characters, and a transition between keys. PS37 introduced a new equation that computes the time it takes to select a word from a list given several factors: scanning and scrolling time, word length, and the index of the selected word in the list.

5.4.4. *Traditional In-Vehicle Information Systems*. Traditional In-Vehicle Information Systems (IVIS) typically consist of a screen surrounded by a series of keys, buttons, and knobs indented to perform tasks such as: turning the radio on,

road navigation, navigating music lists, etc. Of the 54 primary studies, six were categorised as traditional IVIS (PS13, PS15, PS18, PS31, PS42, and PS43). Figure 11 shows how the operators were extended in the new models. The following subsections elaborate further on these operators, heuristics, equations, and metrics.

(1) *Preserved Original Operators*. Four of the five original KLM operators were preserved in the IVIS extensions. PS31 and PS42 document different stages of the same research that enhanced the original KLM for traditional IVIS systems; both studies preserved *K*, *M*, and *P*. One other study, PS43, provides a model for rapid user interface prototyping in the IVIS context and incorporates a modified KLM for that purpose. Of the original operators, only *R* was utilised from the KLM.

(2) *Adapted Original Operators*. *H* was modified from its original values in PS31 and PS42 to reflect new homing interactions between the IVIS and the steering wheel. Similarly, PS15 decomposed *H* into two operators, *Rn* and *Rf*, for reach-near (from the steering wheel to other parts of the wheel) and reach-far (from steering wheel to IVIS). It also presented age-adjusted unit times for older drivers. The study also dissected *K* and *R* into refined operators and replaced the original value of *M* with 1.50 seconds and an age-adjusted value of 2.70 seconds. *M* was also modified by PS43 with two new values based on its placement after *R* and their new turn operator. PS13 adapted *K* for an enter keystroke along with a down keystroke. *K* was also modified by PS43

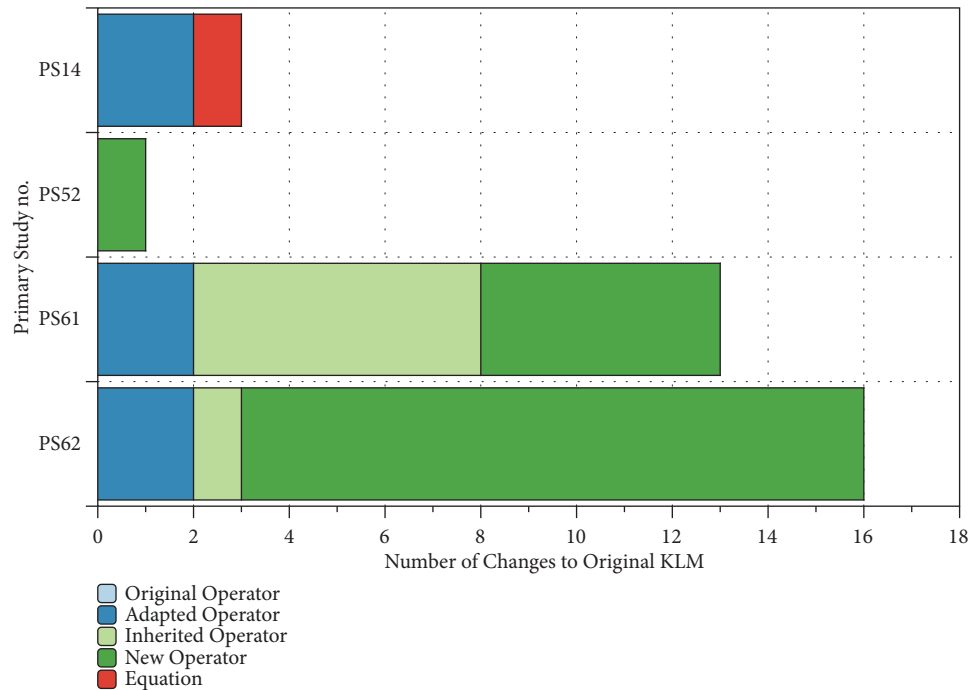


FIGURE 12: The number of original operators preserved, adapted, inherited from other extension or literature, and newly introduced in the touch-based IVIS category.

and PS18 by decomposing the original operator into specified actions.

(3) *Inherited Operators*. PS43 incorporated their extended the KLM with a prototyping model in which three operators were inherited from other extensions: *H* [69], *F* (move finger between controls), and attention shift from PS30. PS15's reach-far operator was inherited by PS31 and PS42.

(4) *New operators*. Only two new operators were introduced in two studies, PS13 and PS43. A reading/decision operator was identified by PS13 to represent the time needed to read an IVIS menu and decide upon actions based on the menu's depth and breadth. A turn operator was introduced in PS43 for tuning a dial (clockwise or counter-clockwise) at various degrees.

(5) *Updated Heuristics*. The placement of *M* was revised to incorporate the turn operator introduced in PS43. The new heuristic dictates that *M* should be placed in two different scenarios with two different values: after *R* and both before and after the user turns a knob.

(6) *New Equations*. The equation to compute the time required to execute a unit task was revised in PS43. Their new equation considered age as a factor as well as visual and non-visual periods that are characteristic of driving and IVIS interactions. PS13 updated the original equation to consider the number of menus encountered and the number of downward scrolls required to read a list item.

(7) *Domain-Specific Metrics*. The occlusion technique is used to simulate common driving distractions that occur when using IVIS systems. In occlusion, users are asked to conduct tasks with an IVIS while wearing computer-controlled goggles that open and shut at regular intervals. This condition imitates the glancing behaviour of drivers who cycle between looking at the IVIS (vision) and driving (non-vision or occlusion) periods. Two metrics are computed by PS31 and PS42:

- (i) Total shutter open time (TSOT): the number of visual periods during occlusion trials with an IVIS.
- (ii) Resumability ratio (R): the degree to which an IVIS task can be performed without looking.

PS31 and PS42's approach to modelling a unit task involved developing the model traditionally using their extended KLM, and then reassessing the sequence of operators by considering the vision/no-vision intervals.

5.4.5. *Touch-Based In-Vehicle Information Systems*. Touch-based IVIS systems feature a touch screen for navigating the IVIS. Of the selected studies, four extended the KLM for touch-based IVIS (PS14, PS52, PS61, and PS62). Figure 12 illustrates the various changes made to extend the KLM. These extensions did not explicitly preserve any of the original operators; thus, the following subsections discuss the adapted and original operators, inherited actions, new operators, and new equations.

(1) *Adapted Original Operators*. PS14 extended the KLM to revise the unit times previously measured in the literature.

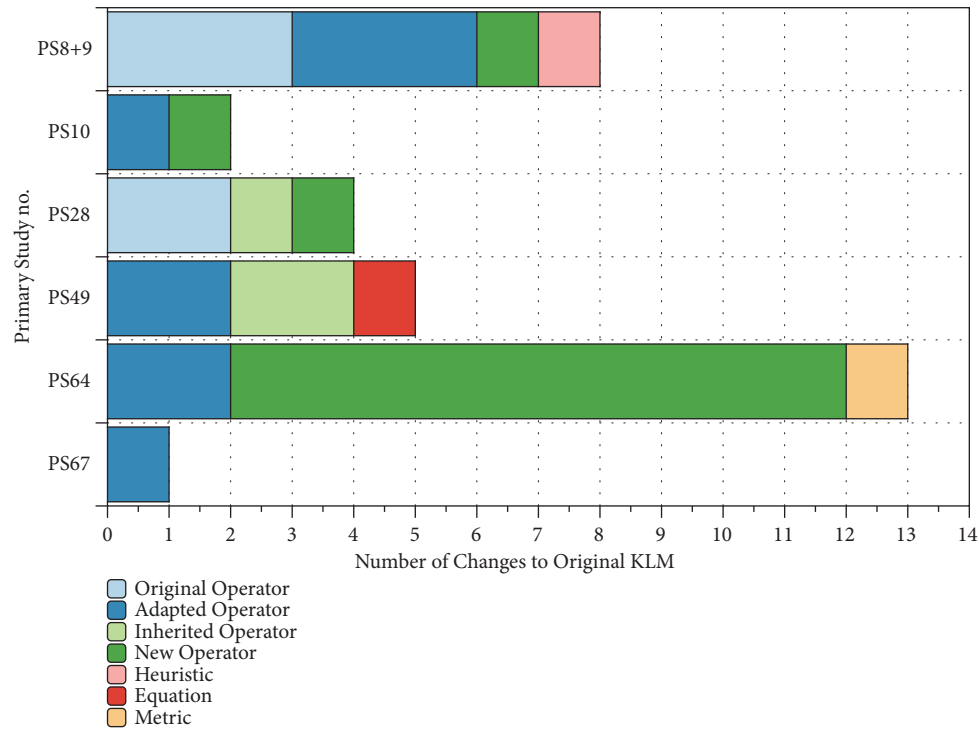


FIGURE 13: The number of original operators preserved, adapted, inherited from other extension or literature, and newly introduced in the specialised setup category.

They argue that the values misrepresented the evaluated IVIS K because the original values were based on a QWERTY keyboard. Therefore, in their study, K and M were revised. For K , several values were considered: letters, numbers, cursor keys, enter, shift, and space. The revision also considered key repetitions. M was adapted to 2.22 seconds in their extension method. K was divided in PS61 to represent function key actions and their repetition. PS61 also considered age-adjusted unit times for these operators. R was adapted by PS62 for wait-while-loading and wait-after-loading, each of which were age adjusted.

(2) *Inherited Operators.* PS15 was revised in PS61 to model interactions with a touch-based IVIS. PS61 inherited and revised the following actions: cursor key pressed once, cursor key after first press, letter key pressed once, letter key after first press, number key pressed once, and number key after first press. The unit times were also adjusted for age. A flicking operator was inherited by PS62 to represent the act of moving a finger in the flick direction (this operator was inherited and revised from PS52).

(3) *New Operators.* PS52 considered the new flick operator in the context of navigating lists of contacts or albums, each of which are age adjusted. Several new operators were introduced by PS61 to re-evaluate the traditional IVIS model, including scrolling through a list, pressing and holding a key, dragging, and first and subsequent slider actions. PS62 developed an extended KLM that overcome a noted shortcoming of the occlusion methods used by PS31 and

PS42. A variety of operators were introduced: flick/scroll return, pressing an on-screen button, quick flick, reach for button, reach for console, read instructions, reposition hand on knob, scroll, search, stop screen, turn knob, and wait for goggles in known and unknown locations to represent the time the user waits for a vision period.

(4) *New Equations.* PS14 determined the retrieval time of a destination from an IVIS, that involved keying in part of the destination name, scrolling through a list of names, or a combination of these approaches. Destination entry tasks were also considered that involved keying in a destination name or a longitude and a latitude. To aid in modelling the KLM extension, the study created a spreadsheet for both tasks in which predicted times were adjusted for age, lighting conditions, and destination. These spreadsheets were used to construct formulas used with equations to calculate the total predicted times for destination retrieval and entry tasks.

5.4.6. *Specialised Setup.* Specialised setups enhance a traditional device with domain-specific controls or involve more than one screen. Seven of the 54 primary studies were categorised as specialised setups (PS8, PS9, PS10, PS28, PS49, PS64, and PS67). Of these, PS8 and PS9 involve the same continuing study. These studies preserved/discarded/adapted original operators, inherited operators from the KLM or its extensions, introduced new operators or equations, updated the heuristics, and identified domain-specific metrics (see Figure 13). The remainder of this section describes the modifications applied to the KLM.

(1) *Preserved Original Operators.* Several operators were preserved from the KLM: *H*, *K*, *P*, and *M*. *H* was utilised by PS8, PS9, and PS28. PS28 also preserved the *K* value of an average non-skilled typist (0.28 seconds). In PS9, *M* was used as is; however, it was not considered in their earlier work (PS8). *P* was utilised from the original KLM in both PS8 and PS9.

(2) *Adapted Original Operators.* The majority of studies in this section adapted operators from the KLM. PS64 developed a GOMS-HRA to dynamically assess the reliability of human operators in nuclear plants. The study introduced two operators, *D_p* and *D_w*, that are analogues of *M* and represent the acts of making a decision based either on an existing procedure or without an existing procedure, respectively. *H* was adapted by PS67 to consider homing actions in hybrid interfaces—particularly for in-air devices such as the Leap Motion sensor. *K* was modified in PS8, PS9, PS10, and PS49. The first two of those studies also divided *R* to identify the time needed to select a new function from a command menu and the time it takes the system to close a polygon in a manual map digitising task. PS49 adapted *P* for keyboard navigation by individuals with motor disabilities. An operator was identified by PS8 and PS9 that adapted *P* to quantify two specialised pointing actions.

(3) *Inherited Operators.* The button click-and-release *BB* operator was inherited by PS49 from PS19, a study that was excluded due to its low score in the review's quality assessment phase. This was also the case for the *M* operator used in PS28 and PS49.

(4) *New Operators.* PS64 extended the KLM to assess the reliability of nuclear plant operators and introduced several new operators: performing a physical action on the control board or in a field, looking up required information on the control board or in a field, obtaining required information on the control board or in a field, producing or receiving verbal or written instructions, and selecting or setting a value on the control boards for fields. A Braille operator was introduced by PS28 to evaluate blind users' interactions during web navigation. A new operator was identified by PS8 and PS9 to represent a button press on a specialised 16-button cursor used for map digitisation.

(5) *Updated Heuristics.* PS8 and PS9's KLM extensions for map digitisation updated the placement of the *M* operator to reflect their modifications to the original KLM operators. They suggested placing an *M* prior to digitising with a snap function before deciding on the next vertex to digitise as well as when deciding whether the digitising task should be ended. For a zooming task, they recommended being careful with the *M* operator because some users may require extra time.

(6) *New Equations.* PS47 provided a basis for an early comparison between keyboard navigation systems (including their newly devised KeySurf system)—particularly when used for tabbing and ID navigation—for people with motor disabilities. PS47 modelled the navigation system using updated

equations that reflected the unique navigation requirements for such systems.

(7) *Domain Specific Metrics.* Human error probability (HEP) was used in PS64 to quantify the KLM operators instead of unit time. Their model approached the KLM differently by arguing for what they declared to be a more important measure of the performance of nuclear plant operators.

5.4.7. *Post-GUI.* In this category, two studies (one of which involves three primary papers) modified the KLM for post-GUI systems. In particular, it extends the KLM for a natural user interface (NUI) (PS57, PS59, and PS68), and for an immersive interface with a projector and mobile navigation (PS66). The rest of this section describes the modified actions, updated heuristics, and new equations in the reviewed studies.

(1) *Preserved Original Operators.* *M* was preserved from the KLM by PS66 for modelling an immersive interface. The extended KLM for a NUI (PS59 and PS68) utilised the original *R*; however, during experimentation this value was ignored.

(2) *Adapted Original Operators.* *D*, while commonly discarded in other setup categories, was adapted by PS57 and modified to reflect drawing gestures in the air, as a user would in a NUI. *M* was adapted by PS59 and PS68, where its values were retrieved from earlier extensions [2, 45]. PS66 modified *P* depending on various user and mobile tracking devices.

(3) *Inherited Operators.* A number of operators were adapted in PS59 and PS68 from prior literature. Two *M* operators (*M_s* and *M_p*) were inherited from MacKenzie [81]. Both operators represent the mental act of preparing to execute subsequent physical actions in response to a stimulus or physical matching event. PS68 inherited *D* from their previous work in PS57. The value of *P* was inherited from Zeng, Hedge, and Guimbretiere [82] in PS59 and PS68 to denote the act of pointing to a target in a NUI.

(4) *New Operators.* Both main studies, as expected, introduced several operators to reflect the new interactions associated with their post-GUI interfaces. PS66's immersive interface required several new operators to represent tasks such as asking questions while using the interface and included start and end of task, question, gap between questions and mentally preparing a response, searching for an answer, reading, and physical movement operators. The NUI KLM also introduced several new operators, some of which were shared in two studies (PS59 and PS68), including holding a hand position, tapping by pushing or moving the hand towards the front, swiping and preparing to swipe, grasping, releasing an open hand, preparing to move the hand from a resting position to the position where a drawing stroke begins, and retracting the hand from the position where the stroke finishes. PS68 later introduced two new operators to reflect the act of pulling and a hand-preference factor.

TABLE II: Summary of statistics used to evaluate the performance of predicted versus observed data.

Computation	Description	Number	Studies
Analysis of variance	The difference between two or more means	1	PS10
Contrast weight analysis	A specific comparison over scores or means.	1	PS34
Correlation analysis	The closeness of the relationship between two or more variables.	11	PS1, PS3, PS13, PS14, PS22, PS31, PS42, PS52, PS57, PS59, PS68
MAPE	Mean absolute percentage error.	1	PS11
Percentage difference	The difference between two values divided by the average of the two values. Shown as a percentage.	1	PS10
Percentage change	The change from one number to another expressed as a percentage increase or decrease.	5	PS8, PS9, PS30, PS46, PS65
Ratio	The ratio between observed and predicted values.	2	PS21, PS66
Regression analysis	The relationship between a dependent variable and one or more independent variables.	1	PS55
RMSPE	The percentage of the square root of the average of the square of all of the errors.	6	PS7, PS43, PS56, PS7, PS59, PS68
T-test	The comparison between the means of two populations to test the difference between the samples.	1	PS24

(5) *New Equations.* The extended model of PS59 and PS68 describes the execution of a NUI task using g-units. G-units are gesture units that identify the time between a hand movement and returning to rest. A single G-unit can contain several gesture phrases (g-phrases) as the hand moves into various position to achieve a stroke. The execution task of the model is the summation of the g-units, each of which is defined in several new equations. PS57 also introduced a new equation from the same study to represent the act of drawing gestures in the air.

(6) *Updated Heuristics.* PS68 updated the original heuristic rules for placing *M*. Rule 0 was updated from the original to consider preparation and *P* operators. Rule 2 was adapted to reflect that when a string of *M*'s belong to a g-phrase, all subsequent *M*s excluding the first one should be deleted. Their updated heuristics also suggest that when a *P* follows a preparation action, then *M* should be deleted (updated from Rule 4). Finally, t new rule was introduced (Rule 5) that stresses that when the model developed is unsure of placement, the number of operators should be emphasised over the placement of *M*s.

5.4.8. *Television.* A single reviewed study (PS47) involved web navigation and text entry (both traditional and predictive) on a television set using a remote control. This study preserved three of the original KLM operators: *K*, *H*, and *M* and considered two different keyboard layouts for

text entry. *P* was adapted to represent the different layouts. A finger movement and a dynamic mental operator were introduced into the extended KLM; the latter considers the additional cognitive load of using a word prediction system. To formulate these text entry tasks, two equations were introduced for the two text entry methods, traditional and predictive, respectively.

5.5. *RQ4: What Was the Research Method Used to Validate a KLM Extension?* The purpose of this research question is to identify the research methods used, if any, to validate the performance of an extended model. The original KLM publication conducted a user study to compare observed data and predicted the KLM's results [2, 6]. The model's performance was evaluated using root-mean-square percentage error (RMSPE), which was calculated as 21%. Of the primary studies, 51.85% (28 studies) conducted user experiments to validate their extended models.

Performance evaluations were commonly statistically analysed using several metrics (excluding PS41 and PS47). Table II summarises the statistics used to evaluate the performance of predicted data versus data observed from users. Correlation analyses were applied in 11 studies (39.29%), while RMSPE was adopted by 6 studies (21.43%). Other statistical measures utilised included contrast weights, mean absolute percentage error (MAPE), percentage difference, percentage change, ratio, regression analysis, and t-tests.

Some studies combined more than one statistical method to confirm their results.

Performance measured via correlation analysis ranged in value from 0.48 to 0.98 among the eleven primary studies. RMSPE values were generally within the suggested KLM bound of 21%, excluding one instance in PS7 where the RMSPE was 31%. The percentage change ranged from -15% to 11% in studies utilising this measure.

6. Discussion

This section summarises the principal findings of this systematic review of KLM extensions. It also addresses the limitations of this review that may threaten its validity. Finally, a discussion of the implications of this review for research and practice is presented.

6.1. Principal Findings. The goal of this systematic review was to examine the purposes for extending the KLM, the methods used to extend the model, how the KLM model was modified, and the techniques used to validate the extended models. The principal findings of this review are as follows:

- (i) This review found diverse studies related to extending the KLM for various domains and device setups. However, the extent to which the KLM was rigorously extended varied based primarily on the purpose of the study.
- (ii) Some studies exhaustively applied research methods for the prime purpose of extending the KLM to new domains or setups or to adapt the models to current situations and technologies. Other studies applied the original KLM to evaluate their applications or devices and included new operators to modify the KLM.
- (iii) Many of the primary studies used controlled experiments to extend the unit times of the KLM or to create new operators.
- (iv) The majority of the studies did not include any type of validation for their extended models. From the studies that did report model validation, controlled experiments were often reported. Performance measures varied; however, the majority utilised correlation analyses, and RMSE (the measure originally used to validate the KLM) was the next most common.
- (v) Only a small number of papers compared the performances of their extended models against the original KLM to determine their effectiveness.
- (vi) The majority of the primary studies were categorised as mobile or tablet, followed by traditional setups and IVIS systems.
- (vii) Several software domains were modelled with extended KLMs; nevertheless, the majority were classified as mobile programmes.
- (viii) *K* and *M* were two of the most commonly preserved and adapted operators, followed by *P*. *D* was almost entirely discarded by most extensions.

- (ix) There is a shortage of studies that address the accessibility needs of disabled users, post-GUI, and Windows-Icons-Menu-Pointer (WIMP) interfaces.
- (x) In the key-based mobile category, half the studies utilised the KLM to calculate text entry with various techniques such as multi-tap or predictive.
- (xi) Two of the selected primary studies substituted the unit times with other measures. PS27 replaced them with scores for each operator and PS64 utilised a domain-specific measure, HEP.

6.2. Limitations. As with other systematic reviews, this review was limited by the search terms and digital databases used. The review was also impacted by selection bias, publication bias, improper or inaccurate data extraction, and data misclassifications. Efforts were taken to alleviate these limitations including the following:

- (i) Setting a wider net with the search terms and digital databases. Database selection was influenced by the inclusivity of the databases, popularity, and recurrences of previous work related to predictive modelling.
- (ii) Publication and selection bias was overcome to some extent by including technical reports and MSc/PhD theses, which comprised the selected primary studies.
- (iii) Data extraction was repeatedly re-evaluated in weekly meetings by the reviewers to guarantee consensus and mitigate inaccurate data extraction and misclassifications.

6.3. Implications. The findings of this systematic review have implications for researchers who plan on refining current extensions or developing new extensions as well as for designers and developers who are considering using the KLM or one of its extensions to evaluate their computer systems.

For researchers, several gaps have been identified in the literatures that lend themselves to future revisions and investigations. Despite the spike in KLM extensions in the past two years (see Figure 3), much of the work done previously requires authentication and revisions for traditional setups and mobile phones. It is unlikely that the unit times measured in the early 2000s would still hold true with current processors and memories. Efforts should be made to re-evaluate useful models with the traditional setups utilised today as well as with mobile phones and tablets that are commonly used.

Tables 9 and 10 summarises the device setups and application domains of the 54 primary studies. While the summaries show a varied selection, several weak areas were identified. Device setups primarily focused on traditional setups, mobile, tablet, and IVIS systems. Despite efforts to develop post-GUI KLM extensions, a shortage still exists in studies that address new setups, including virtual and augmented reality, tangible user interfaces, physical interfaces, tabletops, large touch displays, and malleable interfaces. All of these setups have been in existence for at least a decade and are costly to develop; thus, they would certainly benefit from

TABLE 12: Summary of high-quality studies based on this review that can guide designers and developers to the most appropriate KLM extensions.

Study	Setup	Domain	Extension	Validation	Quality score
PS3	Traditional	Spreadsheet	Experiment	Experiment	10
PS8-9	Specialized	GIS	Observation, experiment	Experiment	8
PS10	Traditional	AAC	Experiment	Experiment	9
PS13	Traditional IVIS	IVIS	Experiment	Experiment	8.5
PS14	Touch-based IVIS	IVIS	Experiment	Experiment	10
PS24	Mobile	Text entry	Literature, experiment	Experiment	8.5
PS30	Key-based mobile	Mobile	Observation, experiment	Experiment	8.5
PS34	Key-based mobile	Voice text entry	Experiment	Experiment	9
PS43	Traditional IVIS	IVIS	Observation, experiment	Experiment	8
PS46	Key-based mobile	Mobile	Observation, experiment	Experiment	9.5
PS55	Touch-based mobile	Custom	Experiment	Experiment	8
PS56	Touch-based mobile	Games	Experiment	Experiment	10
PS68	NUI	Custom	Literature, observation, experiment	Experiment	8.5

predictive models to determine performance in early design phases. While a reasonable array of application domains were investigated, the distribution of studies across these domains was uneven. Concentrated efforts were directed toward mobile applications and IVIS systems, leaving considerable room for further research into domains such as medical IT setups.

Extensions to the KLM commonly occur as a result of experiments to extract new actions and unit times. Figure 6 illustrates the research methods utilised by the reviewed studies to extend the KLM operators, heuristics, and unit times. However, when extending operators and heuristics, the majority of studies did not conduct experiments. While this could be expected for setups similar to the one used to extend and validate the original KLM, it is not ideal for new domains or device setups. Operators determined from normative actions could be useful but may fall short of detecting actions (particularly those relating to mental acts M) that are best observed. It is essential for an appropriate research method to be adopted to develop operators to measure human behaviours. When extending operator unit times, the majority of studies conducted experiments to empirically assign values to their adapted or new operators. This approach is also advisable for future researchers because it ensures accurate and up-to-date measurements. It should also be noted that combinations of research methods strengthen the findings by taking full advantage of their combined benefits.

Of the 54 primary studies selected, only half conducted validation studies to confirm the efficacy of their proposed extended models (see Table 11). At times the same experimental results were used for both extending and validating the models, which clearly lends itself to bias. When a new extension is proposed, it is vital that experiments be conducted to provide empirical evidence of the extension's effectiveness. This calls for more controlled experiments to determine how well the proposed extensions perform. For extensions developed for traditional setups, a comparative assessment against the original KLM could be used to determine how well the new models perform against a stable usability model. Such a comparison might even be possible with setups that rely heavily on the original operators in the KLM.

TABLE 13: General information form.

Question	Answer
Reviewer name	
Primary study ID	
Primary study title	
Author name	
Year of publication	
Publication details	
Type of publication	<input type="radio"/> Conference proceeding <input type="radio"/> Journal article <input type="radio"/> Technical report <input type="radio"/> Thesis (MS or PhD) <input type="radio"/> Extended abstract <input type="radio"/> Other

A further finding was that the majority of reviewed extensions do not provide guidance or suggestions to help designers and developers apply the altered model to their product or computer system. Despite its simplicity, the application of the KLM or any of its extensions requires skill to ensure correct measurements of execution. Several tools (e.g., CogTool) have been developed to automate this process, but these are typically limited to traditional setups. Another observation from the review was that the expert level of users, in the case of most reviewed paper, was not disclosed; the users were merely declared as experts. However, what makes a user an expert? The answer to this question is highly subjective and depends on the perspective of the model developer. This in itself impacts the unit times collected for the operators and thus, the validity of the validation results. For researchers, we find that this issue could be mitigated by a clear definition of expertise that could be consistently applied across domains and device setups.

For designers/developers, we recommend the use of Table 12 to select an appropriate model given their products' domain and device setup. All the studies listed in the table

TABLE 14: Quality assessment form.

Quality Assessment	Mark		
	Yes	Partly	No
The primary study presents a clearly stated purpose for KLM extension	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The primary study extends the KLM with new operators or modifications	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The primary study clearly defines the research method used to extend the KLM	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The extension study's methodology is adequate and repeatable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The extension study clearly states the results and findings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The primary study clearly defines the research method used to validate the extended KLM	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The validation study's methodology is adequate and repeatable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The validation study clearly states the results and findings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The primary study presents a comparative analysis of the extended KLM against the original KLM	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The paper has been cited by other authors and contributes to the literature	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality score			

were ranked as Very High or High during the quality assessment phase and conducted experiments to extend and validate their models. It is also important to compare results from different extensions to determine the one best suited to the target users' actions and perceptions. It should be noted that, at times, the KLM or one of its extensions may be unable to address all the human behaviour anticipated in a product. In this case, combining two or more models is possible but not recommended without thorough investigation.

7. Conclusion and Future Work

KLM is popularly used in the literature to evaluate system design early in the development phase to determine probable performance times for skilled error-free tasks. Over the years, several extensions have been created that modify the original KLM to consider revisions of the original operators, varied device setups, and varied domains. This paper presented a systematic review that summarises the existing KLM extensions developed in the literature. From an initial 2,444 studies, 68 unique publications were selected for the review. Information was extracted from the selected studies, which allowed the reviewers to obtain conclusions to identify common techniques, find research gaps, and construct guidelines.

In future work, we intend to extend this systematic review and plan for future research in various ways:

- (i) Perform a systematic review that addresses the research question "What publications have utilised the KLM or one of its extensions to evaluate the efficiency of their designs and how?" We intend to apply the information obtained from this review.
- (ii) Develop a methodology with a formal protocol for extending the KLM that ensures an exhaustively assessed model.
- (iii) Offer a guide for applying the KLM and its various extensions to guarantee correct application.
- (iv) Review the term "expert" in an attempt to provide a unanimous definition for skilled user behaviour in the KLM and its extensions.

Appendix

A. Quality Assessment and Data Extraction Form

Tables 13, 14, and 15 demonstrate the respective forms used to extract general information, quality assessment, and data from the primary studies.

B. Practitioners Guide to KLM Extensions

Table 12 presents a summary of the high-ranking (based on this review's quality assessment) primary studies that utilised research methods to extend and validate their extensions in various domains and with various device setups. This list can help practitioners determine the best extension to use with their own products.

Conflicts of Interest

The authors declare that they have no relevant or material financial conflicts of interest that relate to the research described in this paper.

Authors' Contributions

Shiroq Al-Megren is an assistant Professor of Information Technology at King Saud University. She holds a Ph.D. in Computing Science from the University of Leeds and an MSc from Newcastle University. She has published numerous research papers on human-computer interaction and participated in several conference and workshops. Her areas of interest include accessibility, tangible user interfaces, touch interaction, and smart distributed systems. **Joharah Saeed Khabti** is a lecturer in the Information Technology Department with a specialisation in Security and Web Development at King Saud University. She is a Ph.D. student holding a Master's degree from George Washington University. Her research interests include image clustering, natural language

TABLE 15: Data extraction form.

RQ#	Research question	Answer
RQ1	What was the purpose of extending the KLM?	
RQ1	What was the intended device and setup?	
RQ1	What was the application domain?	
RQ1	Who were the intended users included in the experiment?	
RQ2	What research method was used to extend the KLM operators or modify heuristics?	<input type="radio"/> Experiment <input type="radio"/> Literature <input type="radio"/> Observation <input type="radio"/> Other
RQ2	What research method was used to extend the KLM operators' unit times?	<input type="radio"/> Experiment <input type="radio"/> Literature <input type="radio"/> Observation <input type="radio"/> Other
RQ2	What were the users' skill levels?	
RQ3	How was the original KLM extended?	<input type="radio"/> KLM operators preserved <input type="radio"/> KLM operators discarded <input type="radio"/> Adapted operators <input type="radio"/> New operators <input type="radio"/> New equations <input type="radio"/> Updated heuristics <input type="radio"/> Application metrics
RQ4	What research method was used to validate the extended KLM?	<input type="radio"/> Experiment <input type="radio"/> Literature <input type="radio"/> Observation <input type="radio"/> Other
RQ4	How was performance measured?	<input type="radio"/> Correlation analysis <input type="radio"/> RMSPE <input type="radio"/> MAPE <input type="radio"/> Other
RQ4	What was the overall performance of the extended KLM?	
RQ4	Was the performance of the extended KLM better than that of the original?	<input type="radio"/> Better <input type="radio"/> Worse <input type="radio"/> Unclear <input type="radio"/> Not measures <input type="radio"/> Other

processing, and human-computer interaction. **Hend S. Al-Khalifa** is a Professor at the Information Technology Department, King Saud University. She has contributed more than 120 research papers to symposiums, workshops, international conferences and journals. Moreover, Professor Hend has served as a program committee member at many national and international conferences and as a reviewer for several journals. Her areas of interest include semantic web technologies, computers for people with special needs, and Arabic NLP.

References

- [1] I. S. MacKenzie, "Motor behaviour models for human-computer interaction," in *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, J. M. Carroll, Ed., pp. 27–54, Morgan Kaufmann, San Francisco, CA, USA, 2003.
- [2] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, USA, 1983.
- [3] B. E. John and D. E. Kieras, "The GOMS family of user interface analysis techniques: Comparison and contrast," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 4, pp. 320–351, 1996.
- [4] B. E. John and D. E. Kieras, "Using GOMS for user interface design and evaluation: which technique?" *ACM Transactions on Computer-Human Interactions (TOCHI)*, vol. 3, no. 4, pp. 287–319.
- [5] R. Malhotra, *Empirical Research in Software Engineering: Concepts, Analysis, and Applications*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2015.
- [6] S. K. Card, T. P. Moran, and A. Newell, "The Keystroke-Level Model for User Performance Time with Interactive Systems," *Communications of the ACM*, vol. 23, no. 7, pp. 396–410, 1980.

- [7] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.
- [8] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Tech. Rep. EBSE-2007-01, Keele University and Durham University, 2007.
- [9] T. L. Roberts and T. P. Moran, "The Evaluation of Text Editors: Methodology and Empirical Results," *Communications of the ACM*, vol. 26, no. 4, pp. 265–283, 1983.
- [10] M. D. Dunlop and M. Montgomery Masters, "Investigating five key predictive text entry with combined distance and keystroke modelling," *Personal and Ubiquitous Computing*, vol. 12, no. 8, pp. 589–598, 2008.
- [11] A. Kankaanpää, "Fids A Flat-Panel Interactive Display System," *IEEE Computer Graphics and Applications*, vol. 8, no. 2, pp. 71–82, 1988.
- [12] A. Kulik, A. Kunert, C. Lux, and B. Fröhlich, "The pie slider: Combining advantages of the real and the virtual space," in *Proceedings of the International Symposium on Smart Graphics*, pp. 93–104, 2009.
- [13] J. R. Olson and E. Nilsen, "Analysis of the Cognition Involved in Spreadsheet Software Interaction," *Human-Computer Interaction*, vol. 3, no. 4, pp. 309–349, 1987.
- [14] H. H. Sad and F. Poirier, "Modeling word selection in predictive text entry," in *International Conference on Human-Computer Interaction*, pp. 725–734, 2009.
- [15] A. D. N. Edwards, "Modelling blind users' interactions with an auditory computer interface," *International Journal of Man-Machine Studies*, vol. 30, no. 5, pp. 575–589, 1989.
- [16] H. Li, Y. Liu, J. Liu, X. Wang, Y. Li, and P.-L. P. Rau, "Extended KLM for mobile phone interaction: A user study result," in *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems, CHI 2010*, pp. 3517–3522, New York, NY, USA, April 2010.
- [17] J. R. Olson and G. M. Olson, "The Growth of Cognitive Modeling in Human-Computer Interaction Since GOMS," *Human-Computer Interaction*, vol. 5, no. 2-3, pp. 221–265, 1990.
- [18] Y. Liu and K. J. Rähkä, "Predicting Chinese text entry speeds on mobile phones," in *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems, CHI 2010*, pp. 2183–2192, USA, April 2010.
- [19] D. M. Lane, H. A. Napier, R. R. Batsell, and J. L. Naman, "Predicting the Skilled Use of Hierarchical Menus With the Keystroke-Level Model," *Human-Computer Interaction*, vol. 8, no. 2, pp. 185–192, 1993.
- [20] R. E. Overill, J. A. M. Silomon, and K. P. Chow, "A complexity based model for quantifying forensic evidential probabilities," in *Proceedings of the 5th International Conference on Availability, Reliability, and Security, ARES 2010*, pp. 671–676, Poland, February 2010.
- [21] Y. Quintana, M. Kamel, and R. McGeachy, "Formal methods for evaluating information retrieval in hypertext systems," in *Proceedings of the 11th Annual International Conference on SIGDOC '93*, pp. 259–272, October 1993.
- [22] M. Schrepp, "GOMS analysis as a tool to investigate the usability of web units for disabled users," *Universal Access in the Information Society*, vol. 9, no. 1, pp. 77–86, 2010.
- [23] P. Haunold and W. Kuhn, "A keystroke level analysis of manual map digitizing," in *Proceedings of the Conference on Spatial Information Theory*, pp. 406–420, 1993.
- [24] M. Pettitt and G. Burnett, "Visual demand evaluation methods for in-vehicle interfaces," *International Journal of Mobile Human Computer Interaction*, vol. 2, no. 4, pp. 45–57, 2010.
- [25] P. Haunold and W. Kuhn, "Keystroke level analysis of a graphics application: manual map digitizing," in *Proceedings of the CHI'94 Conference on Human Factors in Computing Systems*, pp. 337–343, April 1994.
- [26] S. Schneegaß, B. Pflöging, D. Kern, and A. Schmidt, "Support for modeling interaction with automotive user interfaces," in *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2011*, pp. 71–78, December 2011.
- [27] H. H. Koester and S. P. Levine, "Modeling the Speed of Text Entry with a Word Prediction Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 2, no. 3, pp. 177–187, 1994.
- [28] P. Gokarn, S. Lobo, K. Gore, P. Doke, and S. Kimbahune, "KLM operator values for rural mobile phone user," in *Proceedings of the 3rd International Conference on Human Computer Interaction*, pp. 93–96, Bangalore, India, April 2011.
- [29] A. Lee, "Exploring user effort involved in using history tools through MHP / GOMS: Results and experiences," in *Human-Computer Interaction*, pp. 109–114, Chapman & Hall, 1995.
- [30] F. C. Y. Li, R. T. Guy, K. Yatani, and K. N. Truong, "The ILine keyboard: A QWERTY layout in a single line," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST'11*, pp. 461–470, October 2011.
- [31] A. Glenstrup, T. Engell-Nielsen, and J. P. Hansen, "Eye controlled media: Present and future state," Tech. Rep. DK-2100, University of Copenhagen, 1995.
- [32] P. Holleis, M. Scherr, and G. Broll, "A revised mobile klm for interaction with multiple NFC-tags," in *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction*, pp. 204–221, 2011.
- [33] D. Manes, "Evaluation of a driver interface: Effects of control type (knob versus buttons) and menu structure (depth versus breadth)," Tech. Rep. UMTRI-97-42, The University of Michigan Transportation Research Institute, 1997.
- [34] D. Chatterjee, A. Sinha, A. Pal, and A. Basu, "An Iterative Methodology to Improve TV Onscreen Keyboard Layout Design through Evaluation of User Studies," *Advances in Computing*, vol. 2, no. 5, pp. 81–91, 2012.
- [35] D. Manes, "Prediction of destination entry and retrieval times using keystroke-level models," Tech. Rep. UMTRI-96-37, The University of Michigan Transportation Research Institute, 1997.
- [36] B. M. Webster, *Implementing a Quantitative Analysis Design Tool for Future Generation Interfaces [Ph.D. thesis]*, Air-Force Institute of Technology, Wright-Patterson AFB, Dayton, OH, USA, 2012.
- [37] P. Green, "Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, no. 18, pp. 987–991, 1999.
- [38] L. Spalteholz, *KeySurf-A keyboard Web navigation system for persons with disabilities [PhD thesis]*, University of Victoria, Victoria, Canada, 2012.
- [39] M. D. Dunlop and A. Crossan, "Predictive text entry methods for mobile phones," *Personal and Ubiquitous Computing*, vol. 4, no. 2-3, pp. 134–143, 2000.
- [40] K. Jung and J. Jang, "A two-step click interaction for Mobile Internet on smartphone," *Communications in Computer and Information Science*, vol. 373, no. 1, pp. 129–133, 2013.

- [41] O. Bälter, “Keystroke level analysis of email message organization,” in *Proceedings of the the SIGCHI conference*, pp. 105–112, The Hague, The Netherlands, April 2000.
- [42] O. M. C. Rendon, F. Estrada-Solano, and L. Z. Granville, “A mashup ecosystem for network management situations,” in *Proceedings of the 2013 IEEE Global Communications Conference, GLOBECOM 2013*, pp. 2249–2255, USA, December 2013.
- [43] C. Nowakowski, “Navigation system destination entry: The effects of driver workload and input devices, and implications for SAE recommended practice,” Tech. Rep. UMTRI-2000-20, The University of Michigan Transportation Research Institute, 2000.
- [44] T.-P. Kang, B. T.-W. Lin, P. Green, S. Pettinato, and A. Best, “Usability of a prototype generation 4 Hyundai-Kia navigation-radio: Evidence from an occlusion experiment and SAE J2365 and Pettitt’s method calculations,” Tech. Rep. UMTRI-2013-11, The University of Michigan Transportation Research Institute, 2013.
- [45] D. Kieras, *Using the keystroke-level model to estimate execution times*, Working paper, University of Michigan, 2001.
- [46] A. D. Rice and J. W. Lartigue, “Touch-Level Model (TLM): Evolving KLM-GOMS for touchscreen and mobile devices,” in *Proceedings of the 2014 ACM Southeast Regional Conference, ACM SE 2014*, Kennesaw, GA, USA, March 2014.
- [47] M. Maragoudakis and N. K. Tselios, “Improving SMS usability using Bayesian networks,” in *Proceedings of the Hellenic Conference on Artificial Intelligence*, pp. 179–190, 2002.
- [48] F. V. III. Thompson, *Evaluation of a commodity VR interaction device for gestural object manipulation in a three dimensional work environment [PhD thesis]*, Iowa State University, Ames, IA, USA, 2014.
- [49] R. Mori, T. Matsunobe, and T. Yamaoka, “A task operation prediction time computation based on GOMS-KLM improved for the cellular phone and the verification of that validity,” in *Proceedings of the Asian Design International Conference (ADC’03)*, 2003.
- [50] K. El Batran and M. D. Dunlop, “Enhancing KLM (Keystroke-Level Model) to fit touch screen mobile devices,” in *Proceedings of the 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2014*, pp. 283–286, September 2014.
- [51] A. Sears and Y. Zha, “Data Entry for Mobile Devices Using Soft Keyboards: Understanding the Effects of Keyboard Size and User Tasks,” *International Journal of Human-Computer Interaction*, vol. 16, no. 2, pp. 163–184, 2003.
- [52] A. Lee, K. Song, H. B. Ryu, J. Kim, and G. Kwon, “Fingerstroke time estimates for touchscreen-based mobile gaming interaction,” *Human Movement Science*, vol. 44, pp. 211–224, 2015.
- [53] B. E. John, D. D. Salvucci, P. Centgraf, and K. Prevas, “Integrating models and tools in the context of driving and in-vehicle devices,” in *Proceedings of the Sixth International Conference on Cognitive Modeling: ICCCM 2004: Integrating Models*, pp. 130–135, 2004.
- [54] O. Erazo, J. A. Pino, and P. Antunes, “Estimating production time of touchless hand drawing gestures,” in *Human-Computer Interaction*, 569, p. 552, 2015.
- [55] R. Myung, “Keystroke-level analysis of Korean text entry methods on mobile phones,” *International Journal of Human-Computer Studies*, vol. 60, no. 5-6, pp. 545–563, 2004.
- [56] J. A. Palilonis, *Active reading on tablet textbooks [PhD thesis]*, University-Purdue University Indianapolis, Indianapolis, IN, USA, 2015.
- [57] S. Jassar, “The digital pen as a human computer interface,” in *Proceedings of the 2004 IEEE International Symposium on Consumer Electronics - Proceedings*, pp. 629–634, September 2004.
- [58] O. Erazo and J. A. Pino, “Predicting task execution time on Natural User Interfaces based on touchless hand gestures,” in *Proceedings of the 20th ACM International Conference on Intelligent User Interfaces, IUI 2015*, pp. 97–109, April 2015.
- [59] Y. How and M.-Y. Kan, “Optimizing predictive text entry for short message service on mobile phones,” in *Proceedings of the HCII*, vol. 5, 2005.
- [60] R. M. Jones, R. E. Wray, J. Zaiantz, B. Bachelor, and C. Newton, “Using Cognitive Workload Analysis to Predict and Mitigate Workload for Training Simulation,” *Procedia Manufacturing*, vol. 3, pp. 5777–5784, 2015.
- [61] T. Jokela, J. Koivumaa, J. Pirkola, P. Salminen, and N. Kantola, “Methods for quantitative usability requirements: A case study on the development of the user interface of a mobile phone,” *Personal and Ubiquitous Computing*, vol. 10, no. 6, pp. 345–355, 2006.
- [62] P. Green, T.-P. Kang, and B. Lin, “Touch screen task element times for improving SAE recommended practice J2365: First proposal,” Tech. Rep. ATLAS-2015-07, Center for Advancing Transportation Leadership and Safety (ATLAS Center), 2015.
- [63] H. Tonn-Eichstädt, “Measuring website usability for visually impaired people - A modified GOMS analysis,” in *Proceedings of the Eighth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2006*, pp. 55–62, October 2006.
- [64] T. Elwart, P. Green, and L. Brian, “Predicting driver distraction using computed occlusion task times: Estimation of task element times and distributions,” Tech. Rep. ATLAS-2015-01, Center for Advancing Transportation Leadership and Safety (ATLAS Center), 2015.
- [65] K. Hinckley, F. Guimbretiere, P. Baudisch, R. Sarin, M. Agrawala, and E. Cutrell, “The Springboard: Multiple modes in one Spring-loaded control,” in *Proceedings of the CHI 2006: Conference on Human Factors in Computing Systems*, pp. 181–190, April 2006.
- [66] M. T. Burns and F. E. Ritter, “Using naturalistic typing to update architecture typing constants,” in *Proceedings of the In Proceedings of the 14th International Conference on Cognitive Modeling*, pp. 169–174, 2016.
- [67] P. Holleis, F. Otto, H. Hussmann, and A. Schmidt, “Keystroke-level model for advanced mobile phone interaction,” in *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems 2007, CHI 2007*, pp. 1505–1514, May 2007.
- [68] R. L. Boring and M. Rasmussen, “GOMS-HRA: A method for treating subtasks in dynamic human reliability analysis,” in *Proceedings of the 2016 European Safety and Reliability Conference*, pp. 956–963, 2016.
- [69] M. Pettitt, G. Burnett, and A. Stevens, “An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems,” in *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems 2007, CHI 2007*, pp. 1515–1524, USA, May 2007.
- [70] S. Ginn, *CLIsis: An Interface for Visu-ally Impaired Users of Apache Isis Applications [PhD thesis]*, University of Amsterdam, 2016.
- [71] M. Esteves, T. Komischke, S. Zapf, and A. Weiss, “Applied user performance modeling in industry - A case study from medical imaging,” in *Digital Human Modeling*, pp. 576–585, 2007.

- [72] M. Nasirinejad, *Mobile focus+context: Exploring the combination of large overviews with high-resolution details in dynamic, mobile work environments [PhD thesis]*, Dalhousie University, Halifax, Canada.
- [73] A. Hashizume, M. Kurosu, and T. Kaneko, "Multi-window system and the working memory," in *Proceedings of the International Conference on Engineering Psychology and Cognitive Ergonomics*, pp. 297–305, 2007.
- [74] A. Christensen, S. A. Pedersen, P. Bjerre, A. K. Pedersen, and W. Stuerzlinger, "Transition times for manipulation tasks in hybrid interfaces," in *Proceedings of the International Conference on Human-Computer Interaction*, pp. 138–150, 2016.
- [75] A. L. Cox, P. A. Cairns, A. Walton, and S. Lee, "Tlk or txt? Using voice input for SMS composition," *Personal and Ubiquitous Computing*, vol. 12, no. 8, pp. 567–588, 2008.
- [76] E. Moreta, *A predictive model for user performance time with natural user interfaces based on touchless hand gestures [PhD thesis]*, Universidad de Chile, Santiago, Chile, 2016.
- [77] I. S. MacKenzie, A. Sellen, and W. A. S. Buxton, "A comparison of input devices in element pointing and dragging tasks," in *Proceedings of the 1991 SIGCHI Conference on Human Factors in Computing Systems, CHI 1991*, pp. 161–166, May 1991.
- [78] S.-h. Kim, G.-m. Kim, and R. Myung, "Hangul input system's physical interface evaluation model for mobile phone," *Journal of Korean Institute of Industrial Engineers*, vol. 28, no. 2, pp. 193–200, 2002.
- [79] B. E. John and A. Newell, "Cumulating the science of HCI: from s-R compatibility to transcription typing," *ACM SIGCHI Bulletin*, vol. 20, no. SI, pp. 109–114, 1989.
- [80] M. Silfverberg, I. Scott MacKenzie, and P. Korhonen, "Predicting text entry speed on mobile phones," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000*, pp. 9–16, Netherlands, April 2000.
- [81] I. S. MacKenzie, *Human-Computer Interaction: An Empirical Research Perspective*, Elsevier, Waltham, MA, USA, 2013.
- [82] X. Zeng, A. Hedge, and F. Guimbretiere, "Fitts' law in 3D space with coordinated hand movements," in *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, HFES 2012*, pp. 990–994, October 2012.



Hindawi

Submit your manuscripts at
www.hindawi.com

