# A Preliminary Study on Pace Rating Using Video Technology

Peeradaech Suwittayaruk,[1] Dirk Van Goubergen,[1] and Thurmon E. Lockhart[2]

1 Department of Industrial Management, Ghent University, Ghent, Belgium
2 Grado Department of Industrial and System Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

**Abstract**

Pace rating has always been recognized and treated as difficult, subjective, and even controversial. Recently, a new idea of pace rating using video technology by showing work motions on a reference video and an actual (real-life) video simultaneously on a screen has been proposed. By adjusting the speed (and thus the pace) of the reference video, industrial engineers or workers without extensive background in time study can synchronize the motion patterns in both videos, thus quantifying the actual pace of the method under study using their visual sense. However, the impact of motion pattern compatibility between the two stimuli (work motions on the reference video and the actual video) has not been investigated in detail. In this study, we elaborate that motion pattern compatibility between the reference video and the actual video has a significant effect on correct determination of the response (the rated pace) in terms of both the accuracy and the precision. © 2012 Wiley Periodicals, Inc.

**Keywords:** Pace rating; Reference video; Actual video; Decision-making pattern; Motion pattern compatibility

## 1. INTRODUCTION

Many organizations are using time study or work measurement techniques to measure and control the amount of time taken to perform a specific task. The main goal of work measurement is to determine the time required by a well-trained, motivated, and qualified operator to carry out a task at a defined rate of working under standard conditions (Kanawaty, 1992). However, current methods of determining the rate of working are subjective, and this is the most challenging aspect of work measurement (Meyers & Stewart, 2002).

Correspondence to: Peeradaech Suwittayaruk, Department of Industrial Management, Ghent University, Ghent, Belgium. Phone: +32 487 74 27 89; e-mail: peeradaech.suwittayaruk@ugent.be

Received: 21 September 2011; revised 22 March 2012; accepted 22 March 2012

View this article online at wileyonlinelibrary.com/journal/hfm

DOI: 10.1002/hfm.20510

The basic purpose of determining the work rate is to normalize or adjust the mean observed time taken for each work element to be performed in order to develop the normal time.

The quality of the standard time is heavily dependent on the evaluation of the work rate. In this case, the rate of working is evaluated by comparing between a worker's performance and the standard performance of a qualified worker on the same task. As time study or work measurement and especially performance rating is not an exact science, bargaining between management and union still occurs in the real world (Kanawaty, 1992). In practice, ratings with errors within ±5% are considered excellent, while those with errors within ±10% are not excellent but good (Meyers & Stewart, 2002).

According to the "fair day's work for a fair day's pay" philosophy, a required work pace must be consistent and fair to both the company and the employee (Matias, 2001). Based on this philosophy, the well-trained, experienced, and healthy operator is expected to perform prescribed work at a normal pace, which is neither too fast nor too slow with reasonable allowances (Niebel, 1993; Niebel & Freivalds, 2003). Currently, there are two references describing normal pace or 100% pace benchmarks (Barnes, 1980; Groover, 2007; Meyers & Stewart, 2002; Mundel & Danner, 1994; Nadler, 1955; Niebel & Freivalds, 2003; Watmough, 1975): 1) walking at 4.82 km per hr (3 miles per hr), taking 68.58 cm (27 inch) steps, and 2) dealing 52 cards into four equal stacks placed at the corners of a 30.48-cm (1-foot) square in exactly 30 s.

Clearly, the walking reference is compatible with rating lower-extremity motions, while dealing 52 cards is compatible with rating upper-extremity motions. It appears likely that compatibility between reference motions and actual motions is based on the two references associated with upper and lower extremities. Compatibility is defined as "things and beliefs work well together" (Sinclair, 2006, p. 278). Research on stimulus–stimulus compatibility is significant in the case of response selection and execution (Verbruggen, Liefooghe, Notebaert, & Vandierendonck, 2005).

In pace rating, two stimuli and a response are involved: the first stimulus is a reference motion pattern; the second stimulus constitutes an actual task, and the response represents a rated pace. There are two traditional methods of pace rating. In the first method, the work pace is evaluated using the relevant visual cues of the motion of an individual performing some task and recording the observed time using a stopwatch or a personal digital assistant (PDA).

Figure 1 illustrates the first method used by traditional industrial engineers (IEs) and raters who are trained using reference videos of walking, card dealing, and industrial operations, such as SAM (Society for the Advancement of Management) rating films (Das, 1964, 1965, 1988; Reuter, 1977, 1978) or TMI (Tampa Manufacturing Institute) videos (Das, 1993; Das, Smith, & Yeager, 1991; Watmough, 1975). Afterward, when rating the pace of an actual task, the IE must compare the observed pace of working with these reference videos in his or her mind. This visual method has several disadvantages and weak points: it is subjective, not reproducible, and has a low confidence level, and the rater needs a long time (40 or more hours) to learn how to rate in a consistent way (Das, Smith, & Yeager, 1991). Moreover, rationally comparing a stimulus (the tempo of work motions on the reference video) stored in long-term memory with a stimulus (the tempo of work motions on the actual work) in working memory to quantify the response (the rated pace) has never been explored and studied in detail.

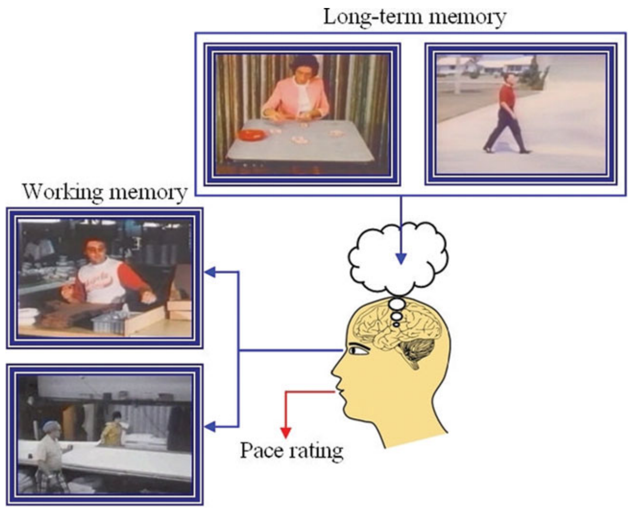


**Figure 1** The traditional method (Watmough, 1975).

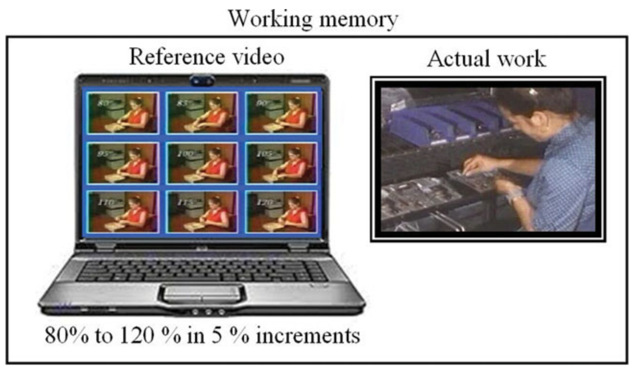


**Figure 2** The second traditional method (Lindenmeyer, 1997; Watmough, 1975).

In the second method, Lindenmeyer (1997) proposed the use of a set of three videos for rating an operator's actual pace of working. Each of the videos contains nine small video images played simultaneously on one screen, showing different paces. The operator's pace can now be matched visually to one of the nine videos on the laptop, in order to determine a rated pace as displayed in Figure 2.

We notice that using these videos for rating can be challenging: the nine stimuli (the nine videos) are rather small compared to the real stimulus under study (the actual workplace), and this can be confusing for the IE who is trying to rate a response (the rated pace) on the shop floor. Also, bringing a laptop to the shop floor might not be suitable in all circumstances.

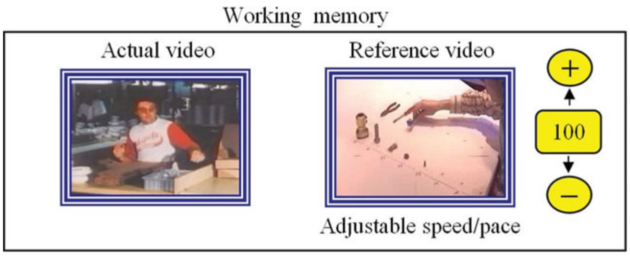


**Figure 3** The currently proposed method (Van Goubergen & Vancauwenberghe, 2006; Watmough, 1975).

Van Goubergen and Vancauwenberghe (2006) proposed new ideas to improve work measurement practices that consist of quantifying waste in work methods and a new proposed method of pace rating, based on video technology, as displayed in Figure 3.

The work motions on a reference video (at pace 100%) and the work motions on an actual video of the work method studied are shown at the same time. As it is extremely difficult to make reference videos at pace 100% that have work motions corresponding to every possible manufacturing task in practice, we choose to define reference work motions that are representative and similar to real activities (reaching, grasping, moving, positioning, and releasing) to produce a reference video, calibrated based on a predetermined time system (PTS) like MTM-1. By adjusting the speed (and thus the pace) of the reference movie, IEs or workers without extensive time study knowledge can synchronize the motion patterns in the two videos and thus quantify the actual pace of the work method under study. Hence, it appears that video technology can help to assess the real pace of a worker in a more rational and reliable way. This is because long-term memory is not involved in the proposed method. Only working memory is addressed in this type of pace rating. By synchronizing the two videos by adjusting the speed of the reference video, additionally, the IE has the ability to control a stimulus (the tempo of the work motions on the reference video) and then focus on comparing only two stimuli (the tempo of the work motions on the reference and actual videos) using vision.

To the best of our knowledge, video referencing method for accurate assessment of pace rating has not been validated. As such, the aim of the present work is to extend the body of knowledge in pace rating with this information. In this study, we hypothesized that compatibility between the work motions on the reference and actual videos will have a significant effect on correctly determining the response (the rated pace).

## 2. METHOD

### 2.1. Participants

Thirty-two participants (graduate Industrial Engineering and Operations Research students at Ghent University, Belgium) were selected as participants. They had backgrounds in pace rating taught in a six-credit work measurement and method engineering course. In this course, after being taught on the concept of normal pace (by showing card dealing and walking videos at pace 100%), students completed pace rating exercises in the traditional way using videos selected from TMI rating series. The actual pace provided together with TMI pace rating video collection was compared with the personally rated values and absolute errors were calculated and analyzed for systematic deviations. Based on the level of information attainment about the deviations, the exercise was repeated on a set of different TMI videos.

In experiments conducted in this article, each participant conducted the experiments by following the experimental procedure as described in Section 2.3.

### 2.2. Apparatus

#### 2.2.1. Pace Rating Software

A software package based on the basic idea of pace rating using video technology (Van Goubergen & Vancauwenberghe, 2006) and developed at the Department of Industrial Management, Ghent University, Belgium, was used to conduct the experiments. Improving the relationship between human and machine system (adapted from Chapanis, 1976) was the initial basis for developing the software (depicted in Figure 4). Transferring information from displays to a sense is the starting point for explaining our model.

Vision, an important human sense, receives the information from the display visualizing the tempo of the work motions on the reference video and the tempo of the work motions on the actual video of the work method being studied, which are shown simultaneously. After the information shown on the display enters the vision, it is stored in the human brain and can then be considered as information processing. Like the information processing models of decision making proposed by Wickens (1992) and Wickens, Lee, Liu, & Becker (2004), when working memory is involved in making a decision, hypothesis generation and selection is the main procedure of information processing.

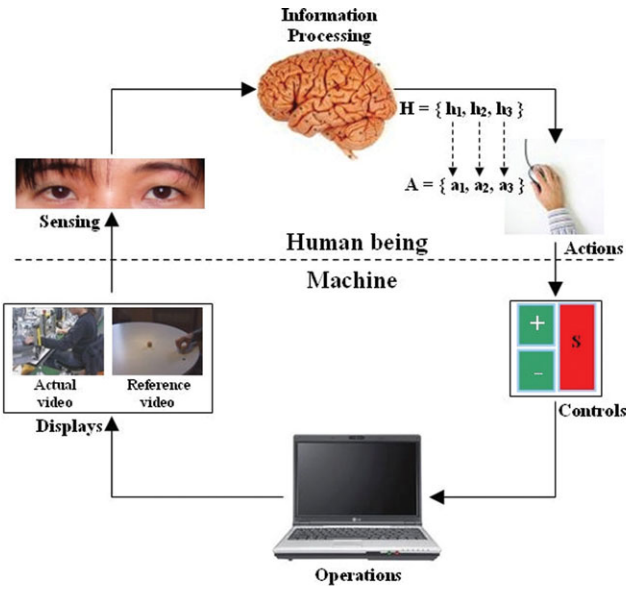


**Figure 4** The human–machine system adapted from Chapanis (1976) for our situation.

In our model, the following possible hypotheses were considered:

$H =$ the set of possible hypotheses
$H = \{h_1, h_2, h_3\}$

where

$h_1 =$ the tempo of the work motions on the reference video is less than that of the actual video
$h_2 =$ the tempo of the work motions on the reference video is more than that of the actual video
$h_3 =$ the tempo of the work motions on the reference video is equal to that of the actual video

After $h_i$ is chosen from $H$ in information processing, the relationship between an action and a control is known. The hypothesis selected can serve as the basis for a course of action. Three buttons: a plus button, a minus button, and a stop button are needed to support alternative actions. We proposed the following controllable choices of alternative actions with regard to the possible hypotheses.

$A =$ the set of alternative actions
$A = \{a_1, a_2, a_3\}$

where

$a_1 =$ clicking the plus button
$a_2 =$ clicking the minus button
$a_3 =$ clicking the stop button

We used JAVA programming to develop our program. The program allowed the speed of the reference movie to be adjusted with 5% increments and 5% decrements. Figure 5 shows a screenshot.

The main goal of the software is to capture the decision-making pattern based on the processing of information. The method of limits was applied as a basic technique to generate the decision-making pattern.

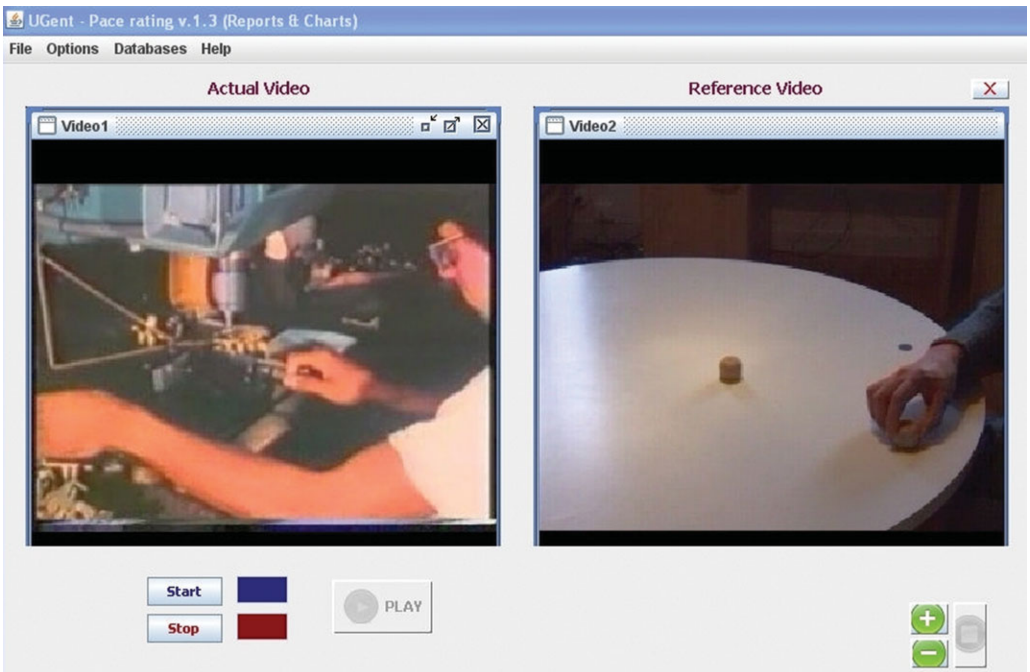


**Figure 5** Screenshot of the software for pace rating.

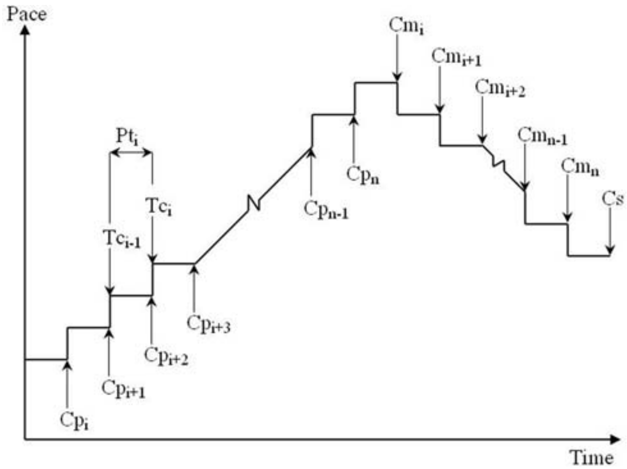


**Figure 6** The basic idea of a button clicked during decision making.

The final response required is $h_3$, which means $a_3$ is activated by the participants. However, the decision to select $a_3$ must continuously be evaluated by comparing the tempo of the work motions on the actual video with the tempo of the work motions on the reference video. In our software, we set the starting point of the tempo of the work motions on the reference video at a minimum tempo difference of 20%. Connecting the possible hypotheses with the alternative actions is the basic idea on which the software is based. The time that elapses before each click (time factor) is recorded in order to compute and visualize the data series of rated paces. Figure 6 displays the basic idea of clicking the mouse during the decision-making process, such as activating the plus button, the minus button, and the stop button. The time is plotted along the $x$-axis and the pace level is indicated on the $y$-axis.

As can be seen from Figure 6, the variables we defined are as follows:

$C\mathrm{p_i}$ = a click on the plus button
$C\mathrm{p_n}$ = the last click on the plus button
$C\mathrm{m_i}$ = a click on the minus button
$C\mathrm{m_n}$ = the last click on the minus button
$C\mathrm{s}$ = the click on the stop button
$P\mathrm{t_i}$ = processing time for each element
$T\mathrm{c_i}$ = time recorded at any click
$P\mathrm{t_i}$ can be calculated as follows:
$P\mathrm{t_i} = T\mathrm{c_i} - T\mathrm{c}_{i-1}$

#### 2.2.2. The Reference Video and Actual Videos

The reference video used in the experiment, presenting an easy motion pattern, was analyzed with MTM-1 and normalized at 100% pace by the authors using Ulead VideoStudio Version 9. The video contains motions, including reaching, (easy) grasping, moving, positioning, and releasing movements with the same distance as shown in Figure 7(a).

Four actual videos were chosen from two different sources to conduct the experiment:

1. Actual video 1, showing an easy motion pattern, was analyzed and developed with MTM-1 by the authors using Ulead VideoStudio Version 9, as shown in Figure 7(b). Like the reference video, the second video also contains reaching, (easy) grasping, moving, positioning, and releasing, but movement distances are different.
2. The set of actual videos 2, 3, and 4 comprises three videos showing hand movements with complex motion patterns rated at 100% pace using the traditional method. These videos were selected from the TMI pace rating video collection (Watmough, 1975): Figure 7(c) shows Film 7, Scene 6 (turn and point flap-shirts); Figure 7(d) shows film H, Scene 6 (heat-seal a part in a bag); and Figure 7(e) shows Film 5, Scene 12 (bar-tack sanitary belts). The 100% pace values have been established by averaging the group opinion of a control group of more than 200 practicing IEs on each scene pictured. These data were provided together with the videos.

### 2.3. Experimental Procedure

A randomized complete block design was designed to conduct the experiments. Each of four sets, consisting of a motion pattern on a reference video and a motion pattern on an actual video, is conducted randomly by each participant (each block). To perform the experiments, the following instructions were given to each participant. Two videos were shown without giving any quantitative information on the pace. Each participant was asked to compare the two videos with regard to pace of working using vision and a hand to click a mouse. As the starting point, the tempo of the work motions on the reference video under study is always shown at a lower pace (at 20%). When the pace is perceived by the participant as being lower, the plus button needs to be clicked by the participant. This causes the tempo of the work motions on the reference video to

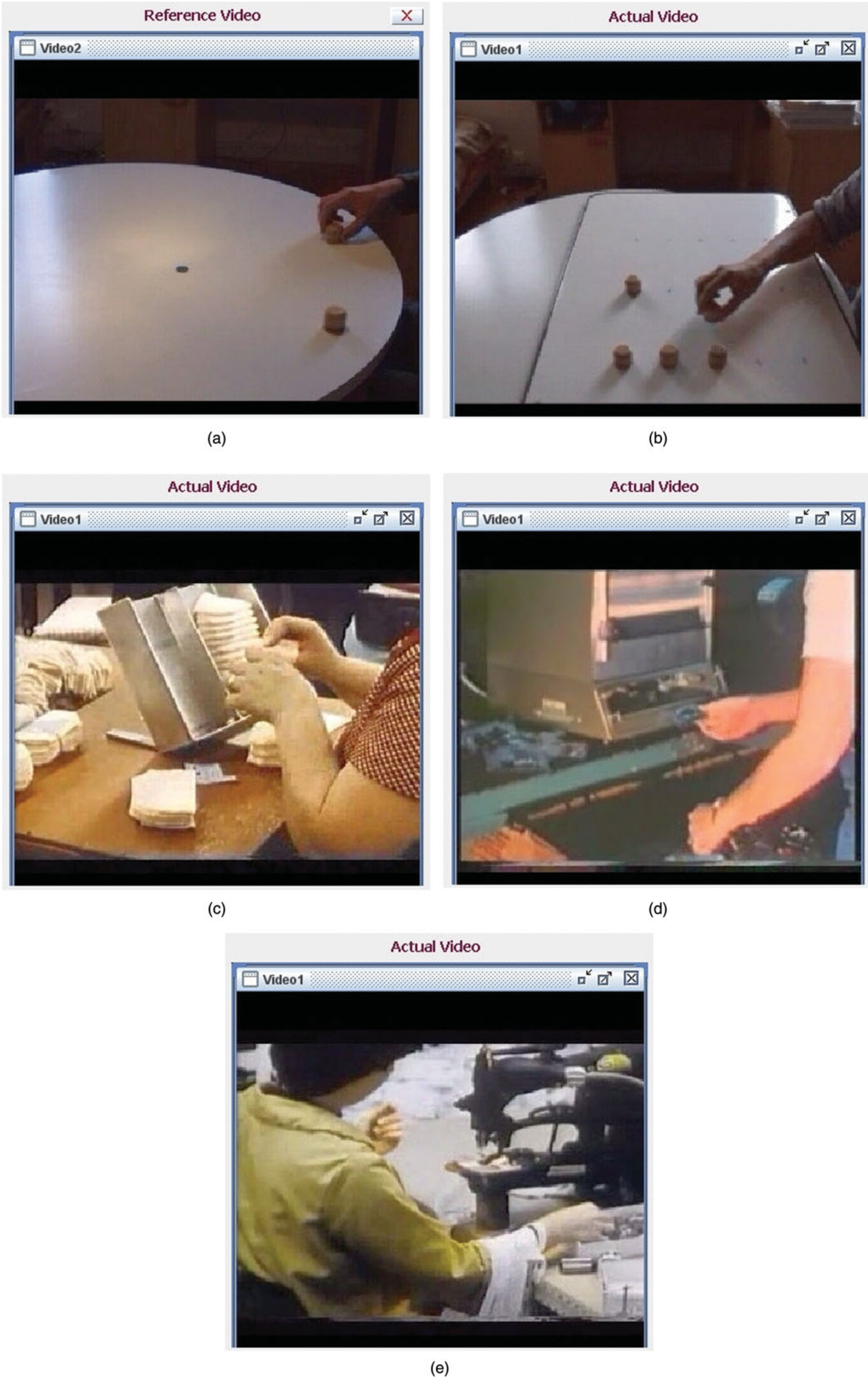


**Figure 7** The reference video and actual videos.

increase. The same question is asked repeatedly until the pace is judged as being higher. Then, the minus button can be clicked until the participant is satisfied with the resemblance. Finally, the stop button needs to be pushed in order to end the experiment. Each click made by the participant while conducting the experiments, along with the corresponding processing time, is recoded in the database and then decision-making patterns are generated. The study was carried out in a room at the Department of Industrial Management, Ghent University, Belgium. The experiments were conducted on an Intel Core 2 Duo 2.20 GHz with 2 GB RAM memory.

## 2.4. Data Analysis

### 2.4.1. Independent Variable

The set, consisting of a motion pattern on a reference video and a motion pattern on an actual video, was defined as the independent variable considered in this investigation. The independent variable was divided into four sets:

1. The easy pattern versus the easy pattern (Figure 8a).
2. The easy pattern versus complex pattern 1 (Figure 8b).
3. The easy pattern versus complex pattern 2 (Figure 8c).
4. The easy pattern versus complex pattern 3 (Figure 8d).

### 2.4.2. Dependent Variables

Two variables, rating accuracy and rating precision, were selected as the dependent variables. The rating accuracy is calculated by Equation 1 below.

$$\text{Average percentage rating error} = \%\overline{e} = \frac{\sum_{i=1}^{n} |\%e_i|}{n} \quad [1]$$

$$\text{where } \%e_i = \frac{rated\ pace - real\ pace^*}{real\ pace^*} \times 100 \quad [2]$$

$n$ = total number of participants involved

*Calculated based on MTM-1 or provided together with TMI pace rating video collection.

The rating precision was computed using Equation 3 below.

$$SD\%e = \sqrt{\frac{\sum_{i=1}^{n} (\%e_i - \%\overline{e})^2}{n-1}} \quad [3]$$

where $SD\%e$ = standard deviation of the percentage rating error

### 2.4.3. Measurement

One hundred and twenty-eight decision-making patterns (32 for each set) were collected. Figure 9 shows an example of a pattern obtained as an output of an experiment. This graph depicts the evolution of the perceived pace as a function of time, showing the information processing of the pace rating of a participant. In this study, however, we used only the end result as the rated pace to investigate the rating accuracy and rating precision. In this example, the rated pace is 95%.

### 2.4.4. Statistical Analyses

For the purpose of investigating the importance of motion pattern compatibility, we first stated the following main hypotheses.

$H_o$: motion pattern compatibility has no significant effect on correct determination of the response (the rated pace).

$H_a$: motion pattern compatibility has a significant effect on correct determination of the response (the rated pace).

As the ability of pace rating depends on how closely the rater is able to evaluate the true or actual pace of the work motions on the actual video, motion pattern compatibility can be determined through two criteria: accuracy and precision.

#### 2.4.4.1. Rating Accuracy

Initially, the null and alternative hypotheses considered were

$H_o$: the four probability distributions of the rating error have the same mean.

$H_a$: at least one distribution of the rating error has a different mean.

A box plot was used to visualize the pace rating data from each set and to verify outliers. To test whether the participants' pace ratings differed significantly,

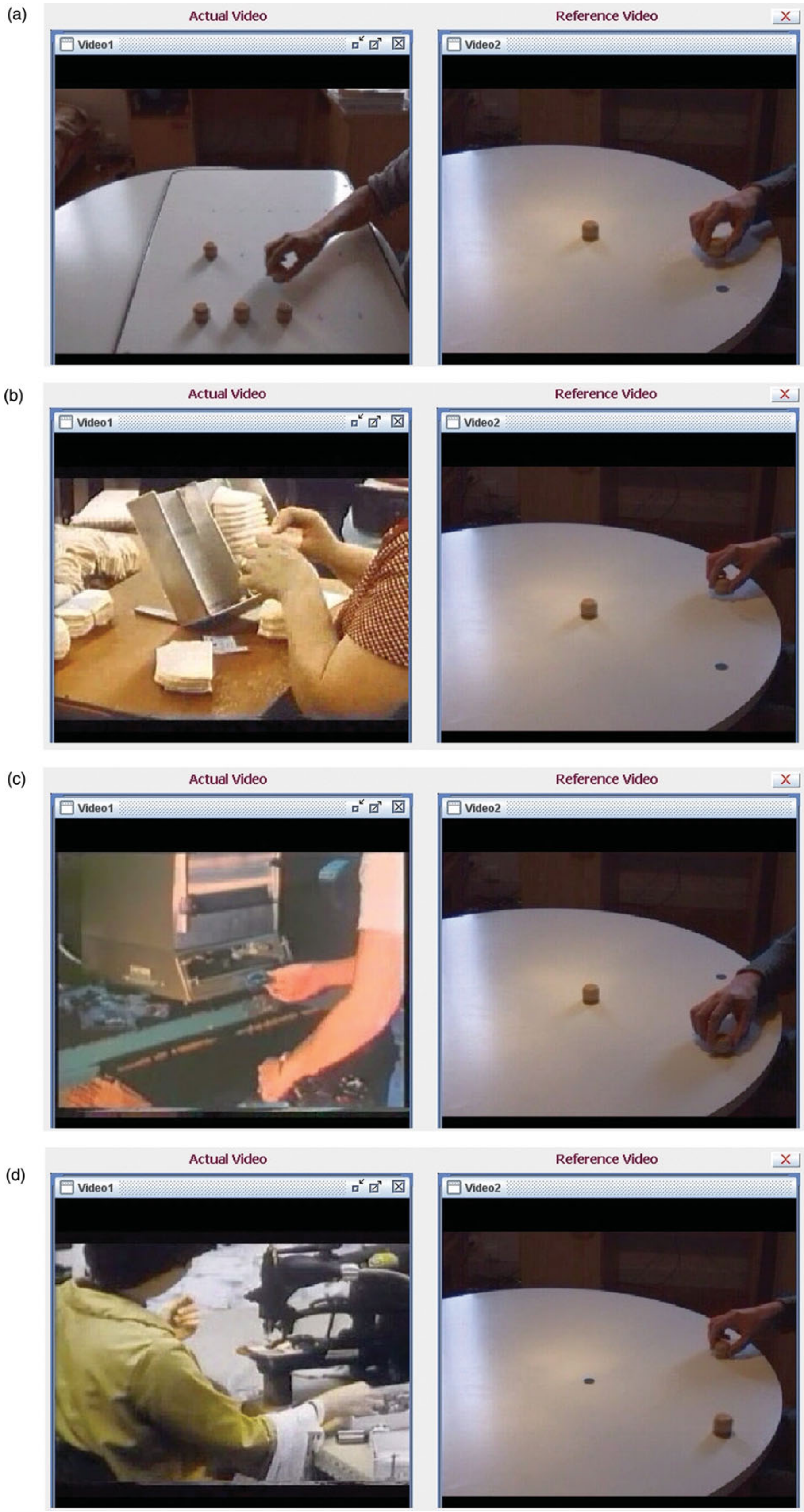


**Figure 8** The set of experiments.

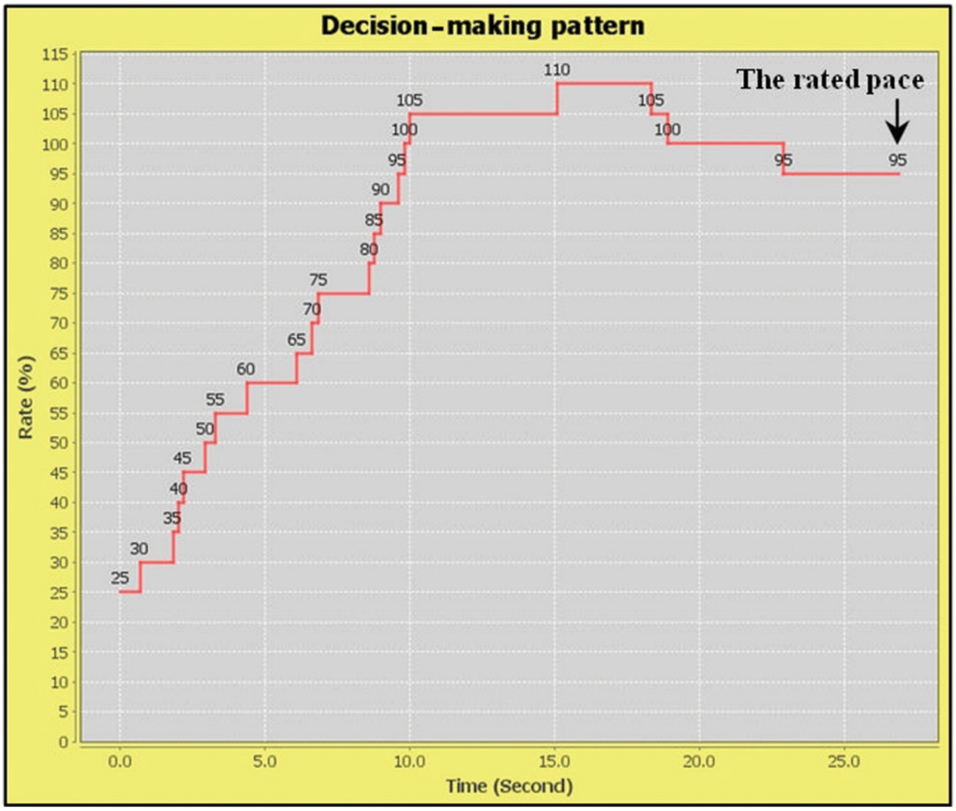


**Figure 9** A decision-making pattern as an output.

analysis of variance (ANOVA) was used. To eliminate any participant's influence, a randomized complete block design was used to conduct the experiments as it reduces and eliminates the variability transmitted from a nuisance factor. In this case, the participant number was the nuisance factor.

We formally tested the normality and equal variance assumptions first because the $F$-test of the ANOVA requires the random variable to be normally distributed with equal variance. To test normality, we used three methods: the Anderson-Darling test, the Ryan-Joiner test, and the Kolmogorov-Smirnov test. Second, to test for equal variances, we used Bartlett's test. Based on the ANOVA, the assumptions were not validated (see the Results section). As an alternative, we tested the difference in population locations (medians) rather than the population means. The Friedman test was chosen as the appropriate nonparametric test to determine whether the population locations (medians) were different. Hence, the adapted hypotheses were set as:

$H_o$: the four probability distributions of the rating error have the same median.

$H_a$: at least one distribution of the rating error has a different median.

A pairwise multiple comparison of sets was performed to determine whether there were differences between pairs of sets by repeatedly applying the Friedman test.

In practice, according to Meyers and Stewart (2002), ratings with errors within $\pm 5\%$ are considered excellent, while those with errors within $\pm 10\%$ are not excellent but good. To define the difference levels of motion pattern compatibility, one sample $t$-test was performed to test the hypotheses for each set of experiments. The null and alternative hypotheses tested in this investigation consist of four sets as shown in Table 1.

#### 2.4.4.2. Rating Precision

The null and alternative hypotheses were

$H_o$: the population variances of experiments in pace rating are equal.

$H_a$: the population variances of experiments in pace rating are not equal.

To analyze the results, we first used Bartlett's test to test the hypotheses. The $F$-test was then performed to characterize the differences between factor-level variances.

**TABLE 1.** Hypothesis Testing on Arranging Motion Pattern Compatibility

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| $H_o : \mu_i = 0\%$<br>$H_a : \mu_i > 0\%$ | $H_o : \mu_i = 5\%$<br>$H_a : \mu_i > 5\%$ | $H_o : \mu_i = 10\%$<br>$H_a : \mu_i > 10\%$ | $H_o : \mu_i = 15\%$<br>$H_a : \mu_i > 15\%$ |

Descriptive and inferential analyses for testing hypotheses were calculated by using the statistical package Minitab Version 13.2. The results of hypothesis testing were considered statistically significant when $p < .05$. To define the difference levels of motion pattern compatibility, we only used levels of significance.

# 3. RESULTS

## 3.1. Rating Accuracy

Figure 10 shows a box plot of the percentage rating error by the different participants for each set. Based on descriptive statistics, no outliers were detected, as illustrated in Figure 10. The graph does show different mean values and variations across the sets and at the same time without overlap. To check whether the differences are statistically significant, ANOVA was performed.

In testing ANOVA assumptions, the $p$-values of Set 1 and Set 2 were less than $\alpha = .05$ for the Anderson-Darling test. There was significant evidence that the two distributions differ from normality. However, the $p$-values of four sets calculated using the Ryan-Joiner test and the Kolmogorov-Smirnov test showed no significant evidence that the four distributions differ from normality. The result of Bartlett's test for equal variances was less than $\alpha = .05$ with $p$-value $= .0001$. Hence, there was compelling evidence that the variances were unequal, so ANOVA could not be used to analyze the data, as the ANOVA assumptions were not satisfied. The Friedman test was then chosen as the appropriate nonparametric test.

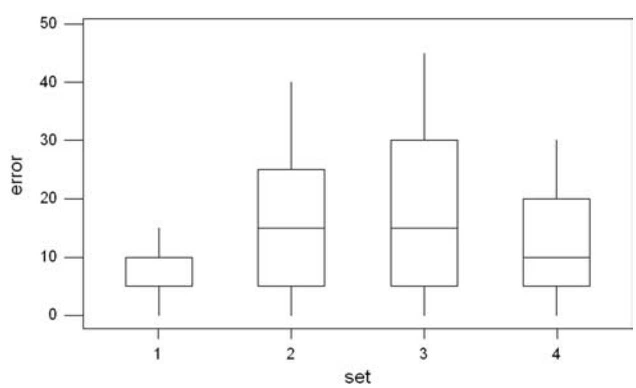


**Figure 10** Box plot of the percentage rating error for each set.

As the $p$-value obtained from the Friedman test taking into account all four sets was equal to .163, the sets are considered not significantly different from one another. Nevertheless, this Friedman test did not show a specific $p$-value for all comparisons of sets. Hence, the Friedman test was performed repeatedly for each pair of sets, which resulted in different $p$-values. To understand clearly the results obtained from these multiple comparisons, Figure 11 shows the different $p$-values obtained in the comparisons of the sets.

Although no significant difference was found between the four sets, the obtained $p$-values could support our hypothesis. In the majority of cases, $p$-values were different, and they can be divided into two groups: those involving Set 1 and those not involving Set 1. The $p$-values of the first group were very close or extremely close to the statistically significant level (.05); especially, comparison of Set 1 with Set 3 gave a value of .059. On the other hand, the results of comparisons not involving Set 1 were extremely far from the $p$-value of .05. Comparison of Set 2 with Set 3 showed, exceptionally, the perfect maximum value of 1.000.

As the Ryan-Joiner test and the Kolmogorov-Smirnov test revealed no significant evidence that the four distributions differ from normality, Figure 12 displays descriptive information of the average percentage rating error and standard deviation of the percentage rating error of each set. The error obtained for Set 1 is less than 10%. On the other hand, the errors for the other sets are more than 10%. The standard deviation of each set was determined. Comparing the results, it was found that, for example, Set 1 gave both the smallest error (8.28%) and smallest standard deviation (4.5), while Set 3 gave the highest error (16.56%) and highest standard deviation (12.73).

To use both the average percentage rating error and standard deviation of the percentage rating error, a parametric test, the $t$-test, was used to test our

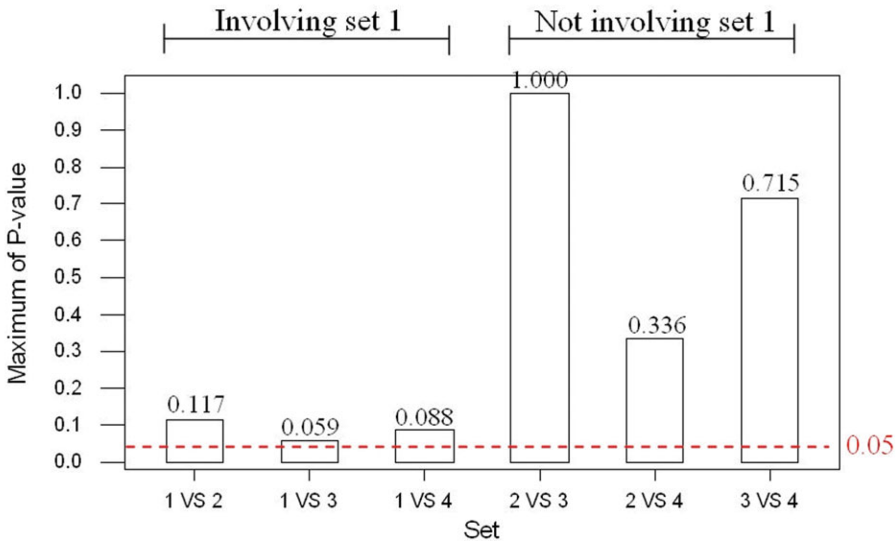


**Figure 11** *p*-values of multiple comparisons.

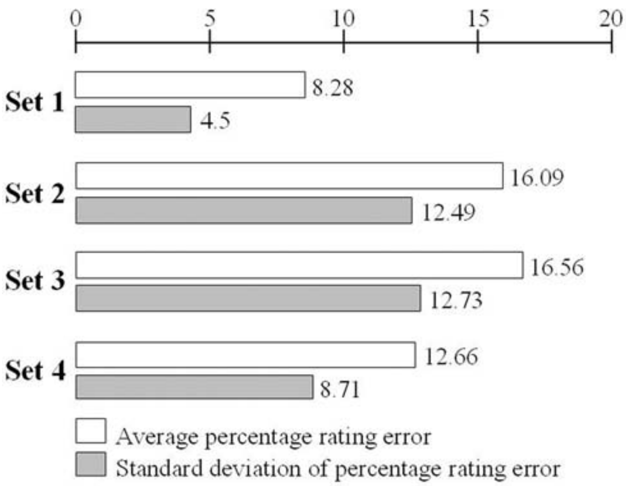


**Figure 12** Average percentage rating error and standard deviation of the percentage rating error.

hypotheses. Table 2 provides the information used to evaluate motion pattern compatibility. For values of $p > .05$ (marked with *), no significant differences were detected at these levels. The *p*-values calculated from the hypothesis testing of $H_o$: $\mu_i = 10\%$; $H_a$: $\mu_i > 10\%$ were chosen as particular data to assess motion pattern compatibility because the initial *p*-value > .05 of Set 1 (.981) was first found by this test. The *p*-value of Set 4 (.047) almost reached the statistically significant level, while those of Set 2 (.005) and Set 3 (.003) were statistically significant with much lower values.

## 3.2. Rating Precision

As can be seen in Figure 13, the *p*-value obtained by Bartlett's test was equal to .000. It is evident that at least 2 standard deviations were significantly different from each other. However, the overall significant *p*-value does not show where the differences lie. In descriptive analyses, clearly, 95% confidence intervals based on standard deviations indicate differences across the sets and at the same time some overlap.

**TABLE 2.** *p*-Values for Defining Difference Levels of Motion Pattern Compatibility

| | $H_o : \mu_i = 0\%$<br>$H_a : \mu_i > 0\%$ | $H_o : \mu_i = 5\%$<br>$H_a : \mu_i > 5\%$ | $H_o : \mu_i = 10\%$<br>$H_a : \mu_i > 10\%$ | $H_o : \mu_i = 15\%$<br>$H_a : \mu_i > 15\%$ |
|---|---|---|---|---|
| Set 1 | 0.000 | 0.0000 | *0.981 (1) | *1.000 |
| Set 2 | 0.000 | 0.0000 | 0.005 (3) | *0.312 |
| Set 3 | 0.000 | 0.0000 | 0.003 (4) | *0.246 |
| Set 4 | 0.000 | 0.0000 | 0.047 (2) | *0.931 |

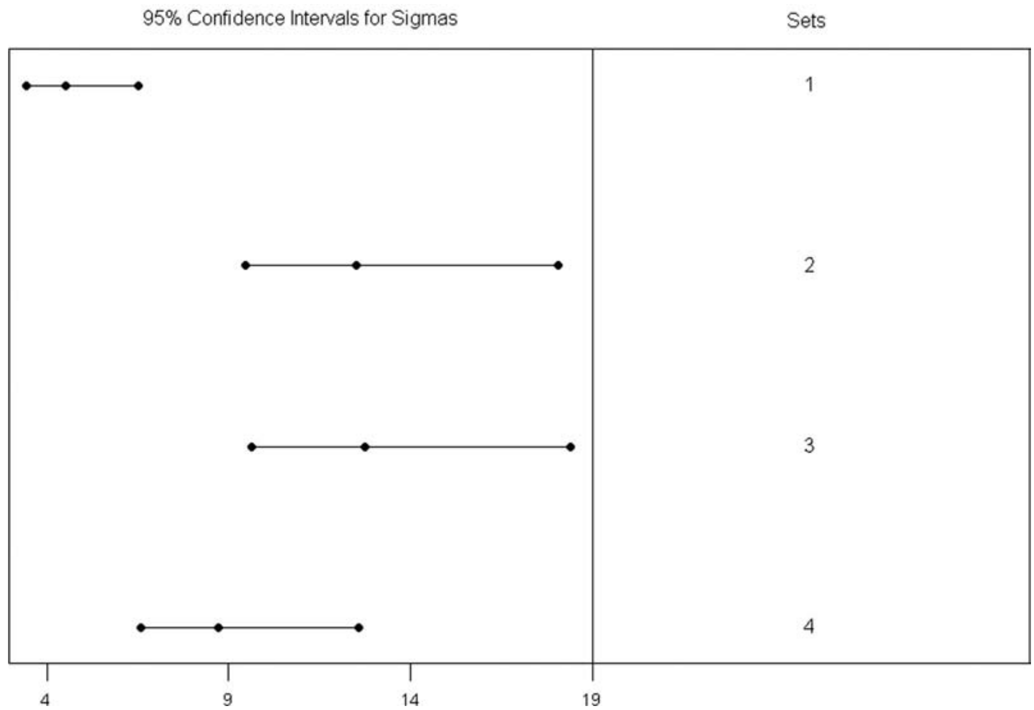


**Figure 13** Test for equal variances for error.

A multiple comparison between sets was performed by applying the $F$-test to determine whether there were differences within pairs of sets. Figure 14 shows the results. There was no significant difference when comparing Set 2 with Set 3, while the rest obtained $p$-values $< .05$. In particular, the results acquired from the comparisons, involving Set 1, had the same $p$-value $= .000$. Interestingly, this result corresponds to what was found for the rating accuracy analysis.

## 4. DISCUSSION

### 4.1. Rating Accuracy

The results of this study support the view that motion pattern compatibility is correlated with the accuracy of pace rating. Despite the obtained $p$-values, the Friedman test, taking into account all four sets, gave .163 as an outcome. However, the results obtained from the multiple comparisons showed different $p$-values. Overall, in comparison, the results not involving the Set 1 pattern reflected considerably higher $p$-values than results involving Set 1. Descriptive data illustrated in Figure 12 show different values for the average and the standard deviation of the rating error of each set. Set 1 (which has an easy motion pattern) has lower values than the other sets showing complex motion patterns (Sets 2, 3, and 4). This is, of course, not a sound statistical conclusion. We used multiple $t$-tests to obtain more evidence. We defined four levels of accuracy.

As stated in the Methods section, according to Meyers and Stewart (2002), rating errors within ±5% are excellent, while those within ±10% are not excellent but good. All $p$-values were .000 for the hypothesis tests of $H_0$: $\mu_i = 0\%$; $H_a$: $\mu_i > 0\%$ and $H_0$: $\mu_i = 5\%$; $H_a$: $\mu_i > 5\%$. This is clear evidence that errors obtained for ratings on the four sets are outside the ±5% range. For Set 1, the initial $p$-value of $.981 > .05$ was only found when testing $H_0$: $\mu_i = 10\%$; $H_a$: $\mu_i > 10\%$. The rest showed $p$-values of less than .05. This means that only the rating of Set 1 (the easy motion pattern) is within the ±10% range. The other sets, containing complex motion patterns, are outside the ±10% range and cannot be considered as good (accurate) ratings. Based on the fourth $t$-test (testing for ±15%), which gave significant high $p$-values, we can conclude the rated paces do stay in the ±15% zone.

Based on our experiments, we can conclude that rating on complex motion patterns gives higher errors than when easy patterns are rated. Hence, motion pattern compatibility has a significant effect on rating accuracy.

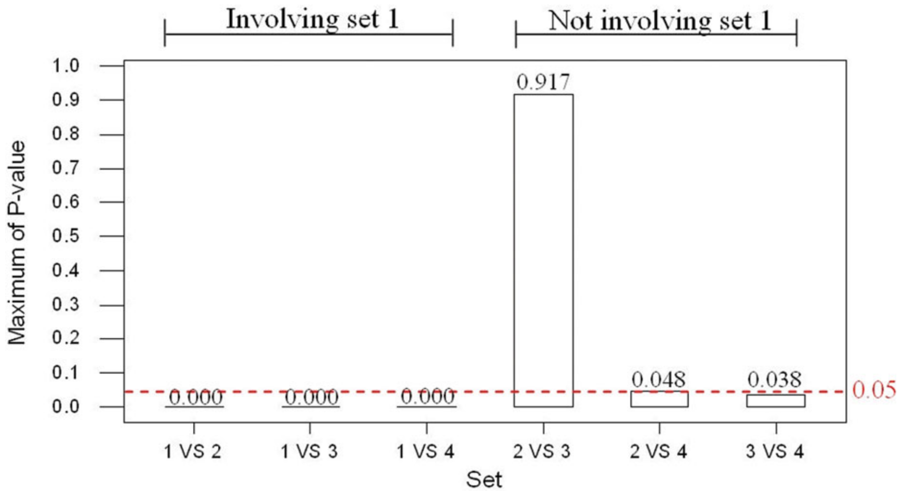


**Figure 14** *F*-test on each set.

## 4.2. Rating Precision

As hypothesized, our experiments prove that motion pattern compatibility has a significant effect on rating precision. The Bartlett's test showed overwhelming evidence, with a $p$-value of .000. Similar to rating accuracy analysis, the $p$-values obtained from the $F$-tests of the multiple comparisons between sets confirmed that our findings were strongly significant. Rating on the easy motion pattern resulted in higher precision than rating on complex motion patterns.

Although no significant difference was found when comparing Sets 2 and 3, this was not the case when comparing Set 2 or 3 with Set 4. For now, it is unclear what is causing this difference. Perhaps some videos are more complex to rate than others, so we might need to introduce a difficulty index. As the information on the actual pace that is provided with the TMI videos (Watmough, 1975) gives only the average group opinion of a control group of more than 200 practicing IEs but no measure of the spread of the sample of data points, the information provided cannot be used to validate this hypothesis. Further investigation is needed to understand what is causing this finding.

# 5. CONCLUSIONS AND FURTHER RESEARCH

The results from this preliminary study points show that compatibility between two stimuli (the work motions on the reference video and the actual video) with regard to motion patterns has a significant effect on correct determination of the response (the rated pace) with regard to accuracy as well as precision. Our current work still has some limitations; for example, some complex motion patterns give better results than others give. For now, we do not have a sound explanation for these differences. An index of difficulty may need to be defined and further research conducted.

Nevertheless, our study contributes to the knowledge on pace rating using video technology and the importance of motion pattern compatibility: comparing videos for pace rating with higher degrees of motion pattern compatibility gives higher accuracy and precision. As such, this paper revitalizes the research on pace rating: it is the first report on pace rating using video technology that indicates the importance of motion patterns compatibility.

Hence, if pace rating based on video technology is used, as proposed here, there is a need for different reference videos with distinct motion patterns. Based on the results of this research, a set of reference videos can be developed, so the IE can select the appropriate one based on motion pattern compatibility and obtain better results for rating the pace of the work under study. Nevertheless, reaching the real pace when rating a complex motion pattern is difficult and challenging, especially in terms of precision. Further research is needed to study the impact of the complexity of the motion patterns.

## References

Barnes, R. M. (1980). Motion and time study design and measurement of work (7th ed.). New York: John Wiley & Sons.

Chapanis, A. (1976). Engineering psychology. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 697–744). Chicago: Rand McNally.

Das, B. (1964). A statistical investigation of the effect of pace and operation on performance rating. International Journal of Production Research, 3(1), 65–71.

Das, B. (1965). Applying programmed learning concepts to instruct in performance rating. Journal of Industrial Engineering, 16(2), 94–100.

Das, B. (1988). The application of DAS's programmed learning method to operator performance rating training in industry. Journal of Methods–Time Measurement, 14(1), 39–43.

Das, B. (1993). Situational factors affecting performance-rating ability. International Journal of Operations and Production Management, 13(3), 49–56.

Das, B., Smith, D. R., & Yeager, R. J. (1991). Personal factors affecting performance rating ability. IIE Transactions, 23(3), 267–272.

Groover, M. P. (2007). Work systems and the methods, measurement, and management of work. Upper Saddle River, NJ: Pearson/Prentice Hall.

Kanawaty, G. (1992). Introduction to work study (4th ed.). Geneva: International Labor Office.

Lindenmeyer, C. (1997). Videos and IE instructional materials [CD-ROM]. C-Four, Inc.

Matias, A. C. (2001). Work measurement: Principles and techniques. In G. Salvendy (Ed.), Handbook of industrial engineering (3rd ed., pp. 1411–1462). New York: John Wiley & Sons.

Meyers, F. E., & Stewart, J. R. (2002). Motion and time study for lean manufacturing (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Mundel, M. E., & Danner, D. L. (1994). Motion and time study: Improving productivity (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Nadler, G. (1955). Motion and time study. New York: McGraw-Hill.

Niebel, B. (1993). Motion and time study (9th ed.). Boston: Irwin.

Niebel, B., & Freivalds, A. (2003). Methods standard and work design (11th ed.). Boston: McGraw-Hill.

Reuter, V. G. (1977, July). The new way to learn pace rating. Industrial Engineering, pp. 16–17.

Reuter, V. G. (1978, May). Learning pace rating. Industrial Engineering, p. 80.

Sinclair J. (2006). Collins COBUILD advanced learner's English dictionary (5th ed.). Glasgow: HarperCollins.

Van Goubergen, D., & Vancauwenberghe, F. (2006). 4 mudas of productivity: Quantifying waste in work May 20–24, 2006. methods [CD-ROM]. IIE Annual Conference, Orlando, Florida.

Verbruggen, F., Liefooghe, B., Notebaert, W., & Vandierendonck, A. (2005). Effects of stimulus–stimulus compatibility and stimulus–response compatibility on response inhibition. Acta Psychologica, 120, 307–326.

Watmough, E. B. (1975). TMI pace rating videos [CD-ROM]. Tempo Manufacturing Institute.

Wickens, C. D. (1992). Engineering psychology and human performance (2nd ed.). New York: HarperCollins.

Wickens, C. D., Lee, J. D., Liu, Y., & Becker, S. E. G. (2004). An introduction to human factor engineering (2nd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.