Proceedings of the ASME 2020 International Design Engineering Technical Conference &
Computers and Information in Engineering Design
IDETC2020
August 16-19, 2020, St. Louis, MO

DETC2020-19136

INDIVIDUAL DIFFERENCES IN DESCRIBING LEVELS OF AUTOMATION

Chase Wentzky
Graduate Research Associate
Mechanical Engineering
Clemson University
Clemson, SC 29634-0921
ecwentz@g.clemson.edu

Joshua D. Summers
Professor
Mechanical Engineering
Clemson University
Clemson, SC 29634-0921
jsummer@clemson.edu

ABSTRACT

Level of automation (LoA) is increasingly recognized as an important principle in improving manufacturing strategies. However, many automation decisions are made without formally assessing LoA and can be made based on a host of organizational factors, like varied mental models used by managers in decision-making. In this study, respondents (N = 186) were asked to watch five different assembly tasks being completed in an automotive manufacturing environment, and then identify "how automated" or "how manual" they perceived the task to be. Responses were given using a visual analogue scale (VAS) and sliding scale, where possible responses ranged from 0 (totally manual) to 100 (totally automated). The activity explored how and when individuals recognized the automated technologies being employed in each task. The tasks of the videos varied primarily by whether the human played active or passive role in the process. Focus group comments collected as a part of the study show how rating patterns revealed functional systems-level thinking and a focus on cognitive automation in manufacturing. While the video ratings generally followed the LoA framework discussed, slight departures in the rating of each video were found.

Keywords: *automation; level of automation; mental models; manufacturing systems design; visual analogue scales*

1 AUTOMATION IN MANUFACTURING

Automation in manufacturing can include robotics, simple mechanization, and information control. This variety of applications can result in many different perceptions of automation. How users understand automation's capabilities and benefits has a marked effect on overall system performance. Although automation in manufacturing has many perceived benefits, not all of these benefits are easily or directly measured. In fact, investing in any advanced technology in manufacturing may have benefits that are not necessarily quantifiable through traditional methods like net present value or return on investment [1]. Further complicating this issue is that technology investment in manufacturing typically has many stakeholders, each with different areas of influence and decision-making power [2]. The decision to automate a particular activity or task can be an intricate and even political process because of competing interests and perceptions of the stakeholders [3].

To address this ambiguity, decision aids for evaluating manufacturing systems have been developed [4–6]. Focusing on identifying optimal configurations of automation, the level of automation (LoA) framework was introduced to aid in systems design and assembly planning [7]. In the manufacturing setting, designing systems with an optimal LoA allows for the ideal distribution of tasks between humans performing manual tasks and machines performing automated functions. Based on their own domain of study, engineers, ergonomists, and other human factors specialists have developed several different schemas for defining the LoA of a system or task. A selection of the definitions for LoA given by the automation literature was gathered in [8] and was updated for this analysis in Table 1.

1.1 LoA Concept in Manufacturing

As shown in Table 1, cross-disciplinary definitions of the LoA share some characteristics (namely, a spectrum describing tasks that range from manual to automatic) but do vary slightly based on their domain and end-use. LoA taxonomies differ in their terminology and granularity, which may make them useful in one setting, but difficult to apply to a broader set of tasks [9,10]. In general, as firms seek to create increasingly advanced production systems, the LoA framework provides a baseline for potential improvement. Several different uses of such a framework have been demonstrated in the manufacturing setting. Applications of the framework range from defining a firm's manufacturing strategy [11,12], optimizing the LoA in assembly [4,7,13], and enhancing other organizational tools to handle increasing levels of automation [6,14].

Table 1. Literature survey of definitions for LoA schemas, adapted from [6].

Ref.	Levels of Automation (LoA) Definition
[15]	The extent to which human energy and control over the production process are replaced by machines.
[16]	The level of automation incorporates the issue of feedback, as well as relative sharing of functions in ten stages.
[17]	The level of automation goes from direct manual control to largely autonomous operation where the human role is minimal.
[18]	The level of automation can be defined as an amount of the manning level with focus around the machines, which can be either manually operated, semi-automated, or fully automated.
[19]	The complimentary degrees to which machines and people make contributions to system processing and output.
[8]	The allocation of physical and cognitive tasks between humans and technology, described as a continuum, ranging from totally manual to totally automatic.
[20]	The amount of automation autonomy and responsibility (highest at the highest level) and the amount of human physical and cognitive activity (highest at the lowest level).
[6]	The degree of automation, the process technology, or to what extent automating using a scale from completely manual (low LoA) to high automated or robotized systems (high LoA).

One widely applied LoA taxonomy for manufacturing is found in [8], and describes seven levels of automation in two domains: "physical" (mechanical and equipment) and "cognitive" (information and control). As identifying cognitive LoA requires significant context about a workstation and its tasks, this analysis will only apply the physical LoA framework in the activity discussed below. The physical levels of automation, as defined by [8], are shown below in Table 2.

Table 2. Levels of physical automation [8].

LoA	Physical Description
1	**Totally manual** - *Totally manual work, no tools are used, only the users own muscle power. E.g. The users own muscle power*
2	**Static hand tool** - *Manual work with support of static tool. E.g. Screwdriver*
3	**Flexible hand tool** - *Manual work with support of flexible tool. E.g. Adjustable spanner*
4	**Automated hand tool** - *Manual work with support of automated tool. E.g. Hydraulic bolt driver*
5	**Static machine/workstation** - *Automatic work by machine that is designed for a specific task. E.g. Lathe*
6	**Flexible machine/workstation** - *Automatic work by machine that can be reconfigured for different tasks. E.g. CNC-machine*
7	**Totally automatic** - *Totally automatic work, the machine solves all deviations or problems that occur by itself. E.g. Autonomous systems*

Traditionally, assigning the physical LoA of a task, or set of tasks, is done through hierarchical task analysis, as demonstrated in [21]. However, providing a numerical value for LoA to an operator or other individual unfamiliar with this taxonomy may not provide much clarity on their role in a human-machine system. Therefore, the objective of this analysis is to understand how individuals perceive the LoA of a task by viewing the task being completed. Further, this study is used to identify what visual criteria may contribute to an individual's mental model of an automated system. These mental models are key to understanding how individuals interact with complex human-machine systems and how individuals interpret current LoA frameworks for manufacturing. The potential consequences of misspecified mental models are of interest [22]. The LoA framework provides the opportunity to assess when stakeholders, when creating mental models of manufacturing systems, may over-specify (see tasks as more automated than they are) or under-specify (see tasks as more manual than they are) their mental models.

One application of understanding these perceptions is in identifying opportunities to adjust LoA as a form of process improvement. This opportunity identification can happen through a variety of methods but is heavily dependent on the mental models regarding automation that managers use to make decisions [23,24] and how one perceives a system's LoA. Based on these perceptions, considerations can be made as to how stakeholders in manufacturing identify certain automation opportunities on the shop-floor, and how to best align those decisions across an organizational structure. Even after an opportunity is identified, those responsible for system maintenance and use are rarely involved in system design. With this in mind, what if the system designer's intended LoA differs from the LoA the user perceives? For this reason, the aim of this study was *not* to evaluate a participant's ability to conduct task analysis. Rather, this work focuses on understanding how respondents perceive automation in each video when shown tasks with varying levels of automation.

1.2 Video Activity Design

To assess manufacturing stakeholders' tendencies while identifying different forms of automation, a series of videos were gathered that would allow individuals to "rate" the LoA in a given system. As LoA is best be described as a "continuum" [8], this analysis was designed to allow individuals to respond on a continuous scale rather than reference a predefined table of discretized automation levels. Accordingly, participants in this experiment were *not* asked to assign a discrete level to each video. Instead, they were asked to score the degree of automation in each video using a visual analogue scale (VAS) ranging from "Totally Manual" to "Totally Automated". This response would then be compared to the relative LoA of each task defined below. Unlike typical Likert-scale formats, the use of a VAS and sliding scale allows participants to describe the tasks in a way that avoids predefining discrete levels [25,26], as discussed in Section 3. As such, responses in this activity can serve as a comparison for how "intuitive" stakeholders find the LoA framework shown above.

The five videos were selected from a montage of assembly tasks posted on the YouTube channel of a large automotive OEM[1] and screenshots of each are shown below. All the videos

[1] Retrieved from https://www.youtube.com/watch?v=adB8xIUTLDI; First accessed *October 10, 2018*

used in the activity are approximately the same length at 30 seconds each and feature one predominant task or the same task repeated multiple times. The tasks occur within the same manufacturing facility and feature assembly tasks relevant to the automotive industry. The following paragraphs include synopses of each video, describes the role of the different agents in each video, and includes the physical LoA as defined by [8], which was assessed by the authors of this paper.

The first task shown in **Video A** (Figure 1) was the application of adhesive to the vehicle chassis in preparation for marriage with the body and lasted 30 seconds. Automated guided vehicles (AGVs) carry the vehicle components in the early phase of assembly into a caged area where two industrial robots apply an adhesive on either side of the body. The video features no human operators involved in the task, with only a few operators present in the background. The task was assessed at LoA = 6 (Flexible machine/workstation) for its use of robotics that could be reconfigured for different tasks.

Figure 1. A screenshot of the task shown in Video A.

Video B (Figure 2) is comprised of operators using impact drills to attach fasteners to a vehicle body in later stages of vehicle assembly and lasted 28 seconds. Operators attach fasteners on multiple parts of the vehicle body but perform an identical task for every fastener. The task contains the highest number of humans present in the video, and prominently features those humans as the main agents of the task. It was assigned LoA = 4 (Automated hand tool) for its use of the hydraulic impact drills present in the video.

Figure 2. A screenshot of the task shown in Video B.

Video C (Figure 3) features adhesive application on the vehicle roof before it is attached to the vehicle body and lasts 31 seconds. An industrial robot holds a vehicle roof and moves it as adhesive is applied along the edge of the roof. Humans are present in the video but are not agents completing any particular task; they are featured close-by to what appears to be the electrical systems governing the robotics. The physical LoA was assessed at LoA = 6, similarly to the first video.

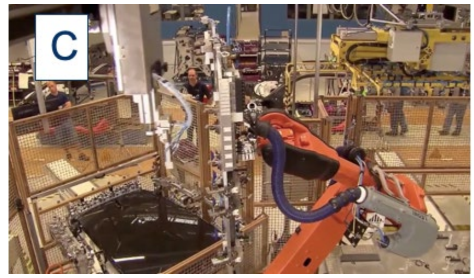

Figure 3. A screenshot of the task shown in Video C.

Video D (Figure 4) shows an operator guiding a vehicle body down to a chassis using a lift assist, where the two parts are joined together and is shown for 30 seconds. The humans in the video are the main agent for completing the task but are clearly assisted by the machinery. This task was assessed at LoA = 5 (Static machine/workstation) because the equipment used is not reconfigurable for multiple tasks.

Figure 4. A screenshot of the task shown in Video D.

The last task shown in **Video E** (Figure 5) is two different operators loading a vehicle roof into a fixture and lasts 32 seconds. After the operators lay the roof in the fixture, one of them presses a button which causes several pneumatic clamps to close down on the roof. Similar to the previous video, the humans have a role in the completion of the task but appear to be assisted by the equipment. It was assessed at LoA = 4 (Level 4: Automated hand tool), because the fixture being used holds the roof in place using pneumatics.

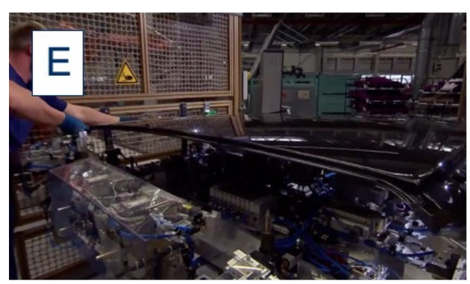

Figure 5. A screenshot of the task shown in Video E.

2 SAMPLE & DATA COLLECTION

The video described above was given to participants in a survey format alongside five scales, so that a single participant would provide five responses – one for each of the five videos. Each participant received the videos in the same order, but the videos were not placed in a way to imply some sequence of tasks. Because of the potential for participants to infer interdependencies between the videos, the order of the five tasks was randomized during the design phase of the video activity. The following section describes the sample of participants given the activity, the variants of the survey, and the protocol used for processing the data before statistical analysis.

2.1 Sample Description

The original sample of participants was made up of 186 individuals, with participants receiving two slightly different variants of the same video activity. The first group completed the video activity using an online survey tool and will be identified in this analysis as the "Computer" group. This group contained 126 participants from a senior-level undergraduate mechanical engineering course taken in Fall 2018. The course focused on design concepts, ideation tools, and systems design. Students were given the link to the activity via email and were directed to a website featuring the same video activity given to participants in the second group. The second group in the sample came from a separate study started in Spring 2019. This group is named the "Traditional" group. It was compromised of 60 individuals, half of which were also undergraduate and graduate mechanical engineering students enrolled in a mechatronics class. The remaining half of this group was made up of 30 individuals employed in the manufacturing sector; they include assembly operators, automation engineers at a large automotive components manufacturer, and designers in two separate automation firms. Unlike the participants in the first group, this portion of the sample completed the activity in a "pen-and-paper" format as part of a larger focus group study. The two variants of the survey activity are discussed in the following section.

2.1.1 Survey Types

The first group of undergraduate students (the Computer group) completed the video activity using an online survey tool. Once the webpage was opened, the students were given the instructions: "First, watch all five tasks shown in the video. Next, watching the video again, rate each task using the scale below each question." Using these instructions, each student could first view all five tasks, then score each task. As shown in Figure 1, the instructions given to rate each task were accompanied by a reminder that participants could re-watch the videos at any time. The scale provided in this variant was in the form of a sliding scale accompanying each question, and would return an integer value between 0 (on the far-left) and 100 (on the far-right) based on where the slider was moved to; it is important to note that this numerical value was not apparent to the participants, and was only returned to researchers on the "back-end" of the survey tool during *post hoc* data analysis.

The Traditional group in this analysis completed the video activity as part of a larger focus group study. In this group, the video activity was completed in-person as part of a "pen-and-paper" survey that started the focus group. Focus groups for this study typically occurred at the participants' place of employment, with the videos shown on a monitor or projector in a conference room. Participants were first instructed to watch all five tasks shown in the video as a group, and were then shown each task again, rating each task as the video progressed during the second viewing. Participants were instructed not to discuss their responses aloud until the entire activity was completed. This mimicked the instructions given to participants in the Computer group, who watched all five tasks, and were then asked to watch each task a second time to rate the perceived LoA. As shown in Figure 1, participants in the Traditional group were instructed to rate the tasks by marking the "100-millimeter line" VAS in exactly one spot.

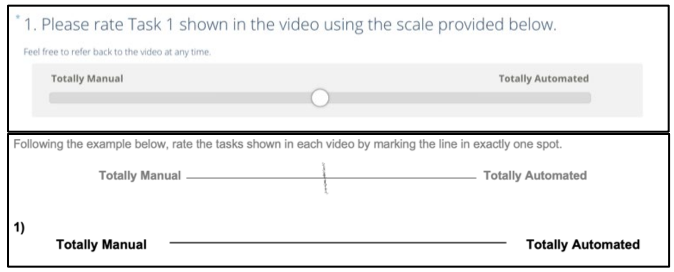

Figure 6. The sliding scale given to the Computer group *(top)*, and VAS given to the Traditional group *(bottom)*.

Like the Computer group, it was not readily apparent to the participants that they were providing a numerical response to each question, but each response was evaluated by researchers after the experiment using an automated 100-millimeter ruler to convert the responses into an integer. As measured by the distance in millimeters from the far-left end of the scale, each response was given a value between 0 and 100. This similarity is one major reason for combining the results collected through the sliding scale and the VAS; they both allow participants to describe their opinions in a seemingly "qualitative" way, while providing relative rankings of the five videos [27]. Although steps were taken during data processing and statistical analysis to mitigate any methodological effects, participant behavior between the two survey types must be considered [28]. Limitations resulting from these differences between the two types will be discussed in Section 6.

2.2 Data Processing

After data collection was completed, the final samples included 126 participants in the Computer group, and 60 participants in the Traditional group. To ensure the validity of the sample collected, data processing considerations were made for each variant of the video activity. The protocol for data processing will be discussed in the following section, and a summary of reasons for sequentially removing samples is shown below in Table 3. It is important to note that if a participant failed on one of the criteria shown below, the entire set of five responses for that participant was excluded from the analysis.

2.2.1 Criteria for Exclusion

The first two criteria in the data processing protocol applied exclusively to the Computer variant of the video activity. The first criterion, "Time too short", refers to the amount of time a

Table 3. A summary of samples removed

		Computer	Traditional
Original Sample		126	60
Reason Removed	Time too short	- 14	N/A
	"50 – 50" Responses	- 4	N/A
	Inverted Scales	- 5	- 1
	"Irregular" Responses	- 3	- 4
Total for Each Version		100	55
Final Total		155	

participant spent completing the activity. The online survey tool used to complete the activity collected timestamps of the start and finish of the activity, with a small group of participants spending less time on the activity than it took to watch the video one time (3 minutes and 12 seconds). These 14 participants' responses were excluded due to the likelihood that the participants did not watch the entire video, and therefore could not give an accurate assessment of the five videos.

The second criterion for exclusion, "50 – 50" responses, applied only to the Computer group was a set of responses that contained multiple instances of "50" as a response. As shown in Figure 1, the default value for the sliding scale was the exact center of the scale, indicating a score of 50. Therefore, scoring a task as 50 may indicate that a participant did not move the slider. Although it may be possible that a participant giving a score of 50 may indicate their actual feelings about the task [28], the four responses with multiple ratings of 50 were considered unlikely to be legitimate and were excluded from the analysis.

2.2.2 Outliers & Irregular Responses

An initial analysis of outliers in the data was completed within responses to individual videos. Outliers were identified using Tukey's method, which defines an outlier as a response one and half times the interquartile range below the first quartile, or one and half times the interquartile range above the third quartile. This approach was considered the most appropriate for this data set, since responses to Videos A and C were highly skewed, and Tukey's method does not assume a normal distribution [29]. While this method helped identify responses that were irregular relative to rest of the ratings given to a single task, simply being identified as an outlier was not a criterion for exclusion from the data. In fact, responding in a unique fashion to a task (e.g. recognizing some form of technology as more automated than another), was an expected outcome. Instead, the outlier analysis was used to identify responses that needed further attention to determine if a response was legitimate.

Using the identification of outliers as a basis, the last two criteria for excluding data applied to both variants of the activity and referred to irregular responses relative to the rest of the data set. There were five responses in which participants seemed to have "inverted the scales" used to indicate whether a task was more manual, or more automated. In other words, these five participants gave extremely low ratings to Videos A and C, and an extremely high score to Video B. Again, while this type of response may indicate the true feelings of participants, responding in an irregular way on three consecutive videos appeared unlikely.

The last seven responses excluded from the data were responses that did not follow the most common pattern of ranking the videos. In the analysis conducted after data collection was complete (discussed in Section 3.2), the responses of the final 162 participants were ranked based on the numerical values derived from their VAS responses. Out of these 162 responses, only seven participants (less than 4%) did not respond to the first three videos in the most common fashion, which brought the final sample for statistical analysis to N = 155; these last seven responses were excluded from the statistical analyses, but will be included in a qualitative discussion in Section 4.2.

3 ANALYSIS & RESULTS

Statistical analysis of the processed data showed a trend among some aspects of the activity, but also identified some unexpected relationships between videos, and differences in systems-level thinking amongst participants. Summary statistics show the relatively strong central tendencies among the first three videos, which are indicative of the common patterns in video ranking shown by the 155 participants. However, inter-item correlations between the five videos showed further relationships between components in each of the videos. To account for differences in behavior between participants in the two survey variants, responses were treated as an ordinal dataset [30].

3.1 Descriptive Statistics

After data processing was complete, researchers were confident in folding together the two variants of the video activity (Computer and Traditional) to bring the total data set to 155 participants and 775 total video responses. To understand relationships between each video, descriptive statistics were generated and are shown below in Table 4.

Table 4. Descriptive statistics of ratings for each activity shown in the video.

Video	*A*	*B*	*C*	*D*	*E*
Mean	95.25	11.8	93.35	56.52	53.83
St. Dev.	7.14	10.93	8.71	16.37	16.97
Median	99	11	98	56	54
Mode	100	0	100	63	54
Range	30	45	38	78	84
Skew	-1.82	0.79	-1.52	-0.18	-0.24
Kurtosis	2.83	-0.02	1.83	-0.51	0.05

Measures of central tendency show the general order of ratings given to each video, with Videos A, B, and C, producing a tightly grouped, albeit highly skewed, distribution of responses. Videos A and C were skewed sharply to the right due to the ceiling effect inherent to using a sliding scale for this type of activity [31], while Video B saw this effect in the opposite direction. Videos D and E showed more spread in their distribution of ratings yet displayed approximately normal distributions. Because the shape of the distribution of responses could not be assumed to be the same for each of the five videos, special care was taken in selecting the appropriate statistical tests when comparing participant responses. Non-parametric tests were employed when analyzing Videos A, B, and C due to their skewed distribution [26]. Likewise, the median and mode of responses to each video were considered for measures of central tendency.

3.2 Video Rankings

To account for the potential effects of survey variant, and the possibility of individual differences in behavior using the VAS, researchers compared the rank order participants gave to each video. For an individual participant, the ratings given to each of the five videos were compared. They were then rank ordered from one (the video given the highest score by that participant) to five (the video given the lowest score by that participant). When two videos were given identical ratings, the system allowed for ties; for example, if Videos A and C were both given a score of 100, they both received a rank of "1", while the video with the next highest rating was given a "3". An example is shown below in Figure 2.

Video	A	B	C	D	E
Rating	97	11	97	29	73
Rank	1	5	1	4	3

Sequence	15143

Figure 7. An example of a sequence generated from the video rankings.

Using this ranking scheme, a pattern emerged in the sequence of video rankings. As shown in Table 5, the most common ranking sequence was rating Videos A and C as either tied, or the first or second highest score (most automated), and Video B as the lowest (most manual) score. For Videos D and E, the likelihood that a participant ranked one video higher than the other was approximately 50%. In other words, while most participants in the study (96%) ranked Videos A, B, and C the same way, the relative rankings of videos D and E was equal to chance. In this sample, 78 participants ranked Video D higher than E, while the remaining 77 did the opposite. The value of these two videos as a discriminant measure between participants will be discussed later in Section 4.2.

Table 5. The distribution of sequences is shown for the sample after data processing.

Sequence	*Count*	*%*
15134	33	21.3%
15243	32	20.6%
15234	24	15.5%
25143	23	14.8%
15143	22	14.2%
25134	21	13.5%
Total	155	

3.3 Inter-item Correlations

To further understand the factors influencing participant rankings of each video, correlations between individual responses were analyzed. Spearman's Rho (r_s) was identified as the most appropriate measure of correlation between the individual ratings given to the videos, rather than Pearson's r. Specifically, the skewed nature of responses to Video A and Video C, along with the possible presence of outliers, made Spearman's Rho a suitable measure for this analysis since it does not assume a normal distribution and uses rank orders [32].

As shown in Table 6, statistically significant values of Spearman's Rho were calculated for every video except E. Of the five significant correlations, some were expected. For example, the value of r_s between Video A and C is 0.591, indicating a strong level of positive association between the responses [33]. As the two videos most prominently featuring robotics, and the highest LoA according to [8], giving a high value to one video predicted a relatively high value for the second. Inter-item correlation also showed a positive relationship in the tendency of participants to rate Video D as "more automated" based on their responses to Videos A and C. However, results from the inter-item correlations show that Video B (the task featuring hydraulic impact drills) was inversely related to Videos A and C. This would indicate that giving a high score (most automated) on one extreme of the scale predicted that the same participants would provide a low score (most manual) on the other end.

Table 6. Inter-item correlations as measured by Spearman's Rho (r_S) values between individual responses to the five activities.

r_S	Video A	Video B	Video C	Video D
Video B	- 0.355^b	--	--	--
Video C	0.591^c	- 0.308^b	--	--
Video D	0.202^a	0.017	0.300^b	--
Video E	- 0.009	0.126	0.093	-0.019

Significant at: ^{a} - $p < 0.05$, ^{b} - $p < 0.01$, ^{c} - $p << 0.001$; N = 155

4 DISCUSSION

Statistical analysis of participant responses showed trends in responses to some of the videos with the possibility for certain video ratings to serve as discriminant measures. Patterns also emerged in individuals' sets of five responses, like the tendencies to group ratings closely together, or to give extreme ratings to certain tasks. Potential reasons for these behaviors, including specific components found in each video, and other qualitative data collected as a part of the focus group study, are discussed.

4.1 LoA Comparisons

As discussed earlier in Section 3.1, a large majority of participants showed a common pattern when rating Videos A, B, and C, with strong grouping around the median rating for each video. This was not a totally unexpected result, given that the LoA as defined by [8] is equal for Videos A and C (Level 6: Flexible machine/workstation). Mann-Whitney tests were used when comparing video responses, with the exception of Videos D and E, since their distributions were approximately normal [34]. For Videos A and C, participants exhibited a slight significant increase in median rating for Video A relative to Video C ($U = 10254, p < 0.05$). This is likely due to the precense of AGVs in the video, or the appearance of the industrial robotics working in concert with the AGVs. In either case, rating Videos A and C as highly automated was also related to rating Video B as highly manual. In a second Mann-Whitney test, it was shown that pariticipants who gave Video B an extremely manual score (defined as a rating of 10 or less) tended to score Video A ($U = 2342, p < 0.05$) three points higher, and Video C ($U = 2299, p < 0.05$) over four points higher, than their counterparts who saw Video B as "more automated".

As shown in Table 7, the median rating and mean ranking of each video is shown. Alongside these participant values are the author's interpretation of "ranking" the videos, with "1" being the most automated and "5" being the most manual.

Table 7. The mean ratings, rankings, author's rankings, and LoA as defined by [8] are shown for each video.

Item	Median Rating	Mean Rank	Author's Rank	LoA (physical)
Video A	99	1.28	1	6
Video B	11	5.00	5	4
Video C	98	1.36	2	6
Video D	56	3.49	3	5
Video E	54	3.50	4	4

It is shown that these median rankings and the author's predicted rankings, correspond closely with the ranking for LoA of each task with the exception of Videos B and E. Video B was given the lowest mean rating of any video by far; Mann-Whitney tests showed a large signifcant difference between Video B, and Videos D ($U = 461, p << 0.001$) and E ($U = 720, p << 0.001$). However, Video B was assigned the same LoA as Video E by researchers in this analysis; in this video, the use of a hydraulic impact drill would indicate an LoA = 4 (Level 4: Automated hand tool) – the lowest level shown in this analysis. Likewise, Video E was considered to have this same LoA = 4. This maybe due to the fact that operators in Video E only place the vehicle roof in a fixture that then uses pneumatics to hold the roof in place. One explanation for this higher score relative to the LoA is that the fixture holding the roof in place appears to contain complicated electronics. However, by assessing only the tasks shown in the 30-second video, it does not appear that these electronics perform any particular task.

4.2 Videos D & E

Due to the common set of responses to the first three videos, ratings for Videos D and E became of special interest as a discriminant measure within the sample of 155 participants. A paired t-test for the difference in means between the two videos showed that the mean ratings were not significantly different ($t = 1.36, df = 154, p > 0.05$), likely due to the large variances found in the sets of responses to each video. However, splitting the sample into two groups using responses to Videos D and E created a nearly "50-50" split, as discussed in Section 3.2. These two groups will be identified as "D-High" (the 78 participants who ranked Video D as third most automated) and "E-High" (the 77 who ranked Video E as third most automated).

4.2.1 Video D Trends

As shown above Table 6, Spearman's correlation (r_s) between videos showed that Video D contained two positive correlations of moderate strength with Video A ($r_s = 0.202$, $p < 0.05$), and Video C ($r_s = 0.300$, $p < 0.01$). These correlations may indicate that participants who rated Videos A and C as being almost totally automated were more likely to "see more" of the automation in Video D. Because the tasks shown in D ("Chassis Marriage") are examples of an automated lift-assist augmenting human ability, this video provides an opportunity for participants to "choose what they want to see" in the video with regards to the LoA. They could choose to see humans as the primary agents completing the task (and rate the task as more manual) or recognize the multiple forms of automation employed in the workstation (and rate the task as more automated). This behavior supports the possibilities discussed in [22] for individual mental models to be misspecified.

This trend is confirmed by Mann-Whitney tests for the difference in medians between the D-High and the E-High groups. Participants in D-High gave Video A slightly higher scores (*2 points)* than those in E-High, although this increase was nonsignificant. For Video C, the increase in median rating between the groups was almost twice that difference (*4 points higher,* $U = 2352, p < 0.05$). These differences in response distribution, which may appear artificially small due to the ceiling effect, reinforce the finding from the inter-item correlations: some individuals were prone to perceive the videos with high LoA as "more automated" than their peers who did not score the remaining videos in-line with the LoA framework. When given a task that appeared more "collaborative" in nature, participants were divided in two approximate halves as to whether they perceived the humans, or the automated equipment, as the predominant agents in the task.

4.2.2 Distance between Videos D & E

The trends described above did not immediately account for the "distance" between ratings for Videos D and E, meaning that whether a participant rated Video D ten points higher than E, or one-hundred points higher, they still belonged to the D-High group. It is important to note that the data gathered from the scales was treated as ordinal, meaning that the *exact* distance between two scores was not of interest. For example, rating one video "20" and a second video "40" did not necessarily indicate that a participant perceived the second task "twice as automated" as the first. Instead, distance between answers was used to describe the likelihood that a participant intentionally ordered one video as higher or lower than another. For the Computer group, the median distance between ratings for Videos D and E was greater than 19 millimeters, and the median distance for the Traditional group was 10.5 millimeters. In total, 70% of participants placed a distance greater than 10 millimeters between the two videos. Because users may find it difficult to describe granular differences using the scales provided in this analysis [28], further investigation was needed.

The first set of observations made about distance between the D-High and E-High groups were their differences in providing "extreme values" to the other three videos. Participants in the Computer group were much more likely give a response of "0" or "100" (32% gave Video B a "0", and 72% gave either Video A or C a "100"), likely due to the mechanics of using an on-screen slider versus using pen-and-paper. Therefore, a definition for an "extreme value" was needed to equate responses to the two survey variants. It was found that participants in the Traditional group responded to Video B with a score of 10 or less 39% of the time and gave ratings of 90 and above to Videos A and C in 68% of cases. These proportions were approximately equal to the proportion of participants who responded with either "0" or "100" in the Computer group, so they became the criteria for an "extreme value" in both variants of the survey.

By grouping the 155 participants by whether or not they gave extremely manual (score of 10 or less) ratings to Video B, or extremely automated (score of 90 or more) ratings to Videos A and C, it was found that ratings near the extremes of the VAS were related to the distance between Videos D and E. The "distance" between Videos D and E is defined as the absolute value of the difference between the two ratings. The value of Spearman's correlation for this "distance" show a moderate positive relationship with the ratings of Videos A and C (r_s = 0.299, $p < 0.01$), and a moderate negative relationship with the ratings of Video B (r_s = -0.263, $p < 0.01$). As shown in Table 8, in each case tested, providing an extreme score to any one of the videos led to at least a roughly 6-point increase in the distance between Videos D and E. This result would support the existence of the trends identified earlier in Section 4.2.1, since it suggests that participants who provided extreme values to the first three videos exhibited differences in behavior when rating Videos D and E. Namely, participants providing extreme scores appeared to have identified differences between Videos D and E that were not identified by their peers (who did not provide extreme scores and essentially viewed Videos D and E as equivalent).

Table 8. Difference of median Mann-Whitney tests are shown for the mean distance between Videos D and E by groups of extreme values.

Gave Video A Extreme High Score?		
	N	Median Distance (mm)
Yes	127	18
No	28	9
Median Difference		+ 9
U = 1199, p = 0.016		
Gave Video B Extreme Low Score?		
	N	Median Distance (mm)
Yes	75	20
No	80	14.5
Median Difference		+ 5.5
U = 2246, p = 0.013		
Gave Video C Extreme High Score?		
	N	Median Distance (mm)
Yes	92	18
No	63	11.5
Median Difference		+ 6.5
U = 2178, p = 0.014		

4.3 Focus Group Comments

In the computer-based version of the activity, no qualitative data was gathered from the participants; however, participants in the Traditional survey group sometimes discussed their individual reasons for scoring each video in a post-activity focus group. This focus group study included manufacturing engineers, management, maintenance technicians, and assembly operators. Focus groups occurred after responses had been entered for each video, and participants were not allowed to change their ratings based on the discussion. Most of the comments fall into two major themes: the tendency to functionally decompose a task into functional inputs and outputs, and the tendency to fixate on the cognitive processes needed to complete each task. Paraphrases of common comments from participants are shown below and shed light on why an individual may respond to the videos in the certain scoring patterns observed.

4.3.1 Functional Decomposition of Tasks

> *"I didn't really say that any of the videos were totally automated, or totally manual."*

The comment that none of the tasks were "totally automated" is reinforced by the view described in [8] that "automation is not all or nothing." This response was common among individuals in the Traditional group, where participants where highly unlikely to mark any task with a score of "0" or "100". Many of the participants in the Traditional group explained that they did not feel comfortable giving these minimum or maximum ratings because they inserted their own fictional intermediate task in the video or viewed the task at a "systems-level". In the case of Video A, many participants felt like they could not say the task was totally automated (a score of "100") because they imagined a human programming the AGVs and industrial robots to work collaboratively. Conversely, in Video B, some of the participants recognized the hydraulic drills as a form of automation and decided to give a score greater than "0". Comments of this type were not limited to any particular occupational group that participated in the study, meaning that this tendency to functionally decompose each video was observed among engineers, students, and operators.

"It looks like that the operators in Video E were loading that roof into a jig for the robot in the third video."

Several participants correctly surmised that the operators loading the roof into the fixture in Video E were doing so in preparation for the task performed by the industrial robot in Video C. This would again imply a recognition of the necessary inputs and potential outputs to each task, which would serve to lower the score of highly automated tasks and raise the ratings of more manual tasks. Participants in this study were not asked about their past experience with hierarchical task analysis, task allocation, or levels of automation, but it is assumed that a large majority had never been formally trained in any of the above skills. Therefore, the idea that some participants would naturally attempt to discretize the workstation into a set of tasks, or imply functional inputs and outputs was an unexpected finding.

"I think those operators in Video B are at a rework station – a machine probably missed those screws."

In regard to Video B, several participants commented that the task shown was a part of a "rework" station. They went on to say that this was probably because a machine had failed to place the fasteners on the vehicle body, and humans were replacing the fasteners that had been missed. Although this comment was not one of the most common, it is similar to the first two comments in that it illustrates the tendency to imply a context of inputs and outputs to each task. This comment only occurred twice during the course of the focus group study, but both instances were during groups with maintenance technicians. This may explain why the participants saw a task that they rated as extremely manual, but implied that it was a task only existed because an automated system had failed. While it is possible that participants in the other groups (made up of mainly manufacturing engineers and students) also recognized this, none of them verbalized this implied context.

4.3.2 Focus on Cognitive Automation

"I was mainly looking to see how attentive the operators had to be in each video. It really depends on what they were thinking about while they were watching it run."

Out of the 162 participant responses that made it into the final dataset, only seven responded in a wholly different way from the larger group (bringing the sample for statistical analysis to N = 155). For those seven participants, the first and third videos were not ranked as the two "most automated" tasks, or they did not rank the second video as the "most manual" task. Although these responses were not included in the statistical analyses shown in Section 3 due to their rarity, the rankings given by these participants could not be totally discounted, as these individuals may have noticed different aspects of the videos than their peers. It is believed that the most likely explanation for these seven irregularities was fixation on the cognitive aspects inherent to the tasks shown.

"They're giving a lot of attention to those drills – it must be difficult to keep them accurate."

This hypothesis is supported by the second common theme for comments in the focus groups: the tendency to perceive the cognitive processes being performed by humans in each video. As discussed in Section 1.1, some LoA frameworks account for both cognitive and physical forms of automation, although the intention of this analysis was only to assess how individuals perceive physical forms of automation. However, participants in this study were not asked to differentiate between these two types of automation; it is possible that participants who gave lower ratings to Videos A and C (which had high levels of physical automation) relative to their peers did so because they fixated on the cognitive performance of the humans present in the video.

Assessing the cognitive LoA of a task requires a large amount of context – much more than the context provided in these videos. However, assigning an exact LoA based on the types of information technology and controls needed for each task was not the aim of this analysis. Rather, these types of comments show that certain participants seemed to fixate on the information technology and controls aspects of the five videos, which affected the perceived degree of automation that they reported. This result supports the conclusions from [8] that a separate cognitive LoA scale is necessary to fully describe automation in manufacturing, even it was a result apparent for all of the participants in this sample.

5 CONCLUSIONS

The findings from this study would suggest that individuals exhibit differences in the visual criteria used to make assessments about the relative levels of automation between manufacturing tasks. Participants in this study responded to each video as to "how automated" or "how manual" they perceived tasks to be, using both a VAS and sliding scale. By using these non-traditional scales, the intent of the study was to evaluate how different manufacturing stakeholders may interpret and rank varying levels of automation. The responses were then compared to a current LoA framework developed for manufacturing, where a majority of participants responded in-line with the seven levels of physical automation defined by [8].

Through careful considerations on how to handle the dataset through non-parametric analysis methods, several trends were identified. It was also shown that certain individuals were likely to gravitate to the extremes of the scales ("Totally Manual" and "Totally Automated"), while others were purposefully reluctant to provide extreme scores. In those cases, participants described the five manufacturing tasks shown in functional terms (having inputs and outputs) and imbuing their own contexts to each task based on their own manufacturing experience. Additionally, some participants recognized aspects of cognitive automation shown in each video, illustrating the increasing complexity of human-machine systems in manufacturing.

The results of this study have potential applications for studying the mental models that managers use in automation decisions, as well as understanding the visual criteria used in identifying automation opportunities on the shop floor. As the set of tasks completed by automation grow more and more complex, understanding how individuals perceive levels of automation will be crucial for system designers and users alike. The tendencies displayed by participants in the study show that individual mental models have the potential to misspecify the degree to which a system is automated; this would mean that manufacturing stakeholders may imply some level of complexity about a system that does not reflect the actual LoA.

6 LIMITATIONS & FUTURE WORK

The current limitations of the experiment mainly revolve around the two variants of the survey given. Most notably, participant behavior near the extremes of the VAS and sliding scale differed among the Traditional and Computer groups. Although the rank-ordering of the video ratings for each participant mitigated some of these effects, alternate methods could have been applied. Additionally, participants in the Traditional group taking the survey in the same room may have biased responses. Discretizing the responses into segments could have also addressed the differences between the two tools, but this would also have introduced difficulties of its own. Further iterations of this study may explore whether or not the findings from this sample are maintained when participants are given a discretized ordinal scale of automation levels instead of a VAS.

Other complications stemming from the use of two survey variants dealt with Computer participants having a referent for the exact middle of the scale – the default rating on the sliding scale for this variant was 50. Scales containing this referent have been shown to result in anchoring effects and variation resulting from education level [31]. This would be especially relevant to this study, since participants in the Traditional group were recruited specifically for their varying work experiences and educational backgrounds. Future studies of this type employing an online survey may seek to use scales more closely mirroring the "pen-and-paper" VAS. Similarly, despite the use of rank-ordering to highlight the difference between Videos D and E, limitations arise from having two videos with relatively low LoA side-by-side. Randomizing video order for each participant in the future may help to mitigate these effects.

Lastly, future versions of this study could give the video activity to participants individually, rather than giving the activity in groups. In this experimental setting, participants could describe the reasoning behind their responses using the "think-aloud" method. Alternatively, each participant could provide feedback regarding their responses individually, instead of in the focus group. Despite the efforts of researchers to encourage each participant in the focus groups to "speak up", it is possible that some participants did not feel comfortable explaining their answers in the group setting [35]. Another potential advantage of conducting the experiment in the individual setting stems from the many participants in the Traditional group who verbally confirmed that fixation played a role in their response patterns. Researchers could employ individual eye-tracking in future studies as a possible method for confirming this fixation on components in each video.

7 ACKNOWLEDGEMENTS

The authors would like to thank the students of Clemson University's Capstone Design course for their participation in this study. They would also like to thank the Advanced Robotics for Manufacturing Institute and their partner companies for their valuable feedback during focus groups in this activity, and the National Science Foundation for their support under grant number CMMI-1463370. The views expressed in this paper are solely those of the authors.

8 REFERENCES

[1] Nixon B., 1995, "Technology investment and management accounting practice," *Br. J. Manag.*, **6**(4), pp. 271–288.

[2] Jain B., Adil G. K., and Ananthakumar U., 2014, "Development of questionnaire to assess manufacturing capability along different decision areas," *Int. J. Adv. Manuf. Technol.*, **71**(9–12), pp. 2091–2105.

[3] Hofmann C., and Orr S., 2005, "Advanced manufacturing technology adoption — the German experience," *Technovation*, **25**, pp. 711–724.

[4] Fasth Å., 2011, "Comparing methods for redesigning, measuring and analysing production systems," *Proceedings of the 4th Swedish Production Symposium*.

[5] Salmi A., David P., Blanco E., and Summers J. D. D., 2016, "A review of cost estimation models for determining assembly automation level," *Comput. Ind. Eng.*, **98**, pp. 246–259.

[6] Salmi A., Dhulia J., Summers J. D., David P., and Blanco E., 2015, "Methods for Selecting Level of Automation: A Critical Comparison of Approaches and Integrated Proposal," *ASME 2015 Design Engineering Technical Conferences and Computers and Information in Engineering Conference-IDETC/CIE 2015*.

[7] Fast-Berglund Å., and Stahre J., 2010, "Concept model towards optimising Levels of Automation (LoA) in assembly systems," *Proc. 3rd CIRP Conf. Assem. Technol. Syst.*

[8] Frohm J., Lindström V., Winroth M., and Stahre J., 2008, "Levels of automation in manufacturing," *Ergonomia*, **30**(3).

[9] Kaber D. B., 2018, "Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation," *J. Cogn. Eng. Decis. Mak.*, **12**(1), pp. 7–24.

[10] Burns C. M., 2018, "Automation and the Human Factors Race to Catch Up," *J. Cogn. Eng. Decis. Mak.*, **12**(1), pp. 83–85.

[11] Säfsten K., Winroth M., and Stahre J., 2007, "The content and process of automation strategies," *Int. J. Prod. Econ.*, **110**(1–2), pp. 25–38.

[12] Winroth M., Säfsten K., Lindström V., Frohm J., and Stahre J., 2006, "Automation Strategies-Refinement of Manufacturing Strategy Content," *17th Annual POM Conference,*.

[13] Romero D., Bernus P., Noran O., Stahre J., and Fast-

Berglund Å., 2016, “The operator 4.0: human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems,” *IFIP International Conference on Advances in Production Management Systems*, pp. 677–686.

[14] Salmi A., David P., Blanco E., and Summers J., 2016, “Standardized Vocabularies for Assembly Systems Modelling and Automation Alternatives Description,” *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE 2016*.

[15] Amber G. H., and Amber P. S., 1962, *Anatomy of automation*, Prentice-Hall.

[16] Sheridan T. B., 1980, “Computer control and human alienation,” *Technol. Rev.*, **83**(1), pp. 60–73.

[17] Billings C. E., 1997, *Aviation Automation: The Search for a Human-Centered Approach*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

[18] Groover M. P., 2002, “Automation, Production Systems, and Computer-integrated Manufacturing,” *Assem. Autom.*, **22**(3), pp. 298–299.

[19] Dekker S. W. A., and Woods D. D., 2002, “MABA-MABA or abracadabra? Progress on human--automation co-ordination,” *Cogn. Technol. Work*, **4**(4), pp. 240–244.

[20] Onnasch L., Wickens C. D., Li H., and Manzey D., 2014, “Human performance consequences of stages and levels of automation: An integrated meta-analysis,” *Hum. Factors*, **56**(3), pp. 476–488.

[21] Mattsson S., Fässberg T., Stahre J., and Fasth, 2011, “Measuring interaction using levels of automation over time,” *21st International Conference on Production Research: Innovation in Product and Production, ICPR 2011 - Conference Proceedings*.

[22] Martignoni D., Menon A., and Siggelkow N., 2016, “Consequences of misspecified mental models: Contrasting effects and the role of cognitive fit,” *Strateg. Manag. J.*, **37**(13), pp. 2545–2568.

[23] Gleeson F., Goodman L., Hargaden V., and Coughlan P., 2017, “Improving worker productivity in advanced manufacturing environments,” *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 297–304.

[24] John C. H. S., Cannon A. R., and Pouder R. W., 2001, “Change drivers in the new millennium: implications for manufacturing strategy research,” *J. Oper. Manag.*, **19**(2), pp. 143–160.

[25] Bishop P. A., and Herron R. L., 2015, “Use and Misuse of the Likert Item Responses and Other Ordinal Measures,” *Int. J. Exerc. Sci.*, **8**(3), pp. 297–302.

[26] Heller G. Z., Manuguerra M., and Chow R., 2016, “How to analyze the Visual Analogue Scale: Myths, truths and clinical relevance,” *Scand. J. pain*, **13**(1), pp. 67–75.

[27] Reips U.-D., and Funke F., 2008, “Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator,” *Behav. Res. Methods*, **40**(3), pp. 699–704.

[28] Couper M. P., Tourangeau R., Conrad F. G., and Singer E., 2006, “Evaluating the effectiveness of visual analog scales: A web experiment,” *Soc. Sci. Comput. Rev.*, **24**(2), pp. 227–245.

[29] Mosteller F., and Tukey J. W., 1977, “Data analysis and regression: a second course in statistics.,” *Addison-Wesley Ser. Behav. Sci. Quant. Methods*.

[30] Kersten P., Küçükdeveci A. A., and Tennant A., 2012, “The use of the Visual Analogue Scale (VAS) in rehabilitation outcomes.,” *J. Rehabil. Med.*, **44**(7), p. 609.

[31] Funke F., Reips U.-D., and Thomas R. K., 2011, “Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level,” *Soc. Sci. Comput. Rev.*, **29**(2), pp. 221–231.

[32] Caruso J. C., and Cliff N., 1997, “Empirical size, coverage, and power of confidence intervals for Spearman’s rho,” *Educ. Psychol. Meas.*, **57**(4), pp. 637–654.

[33] Chen P. Y., Smithson M., and Popovich P. M., 2002, *Correlation: Parametric and nonparametric measures*, Sage.

[34] McKnight P. E., and Najab J., 2010, “Mann-Whitney U-test,” *Corsini Encycl. Psychol.*, p. 1.

[35] Harrell M., and Bradley M., 2009, *Data Collection Methods Semi-Structured Interviews and Focus Groups*.

