Developing a Maturity Model for Information Security Awareness Using a Polytomous Extension of the Rasch Model

Tobias Fertig, Andreas E. Schütz, and Kristin Weber
Faculty of Computer Science and Business Information Systems
University of Applied Sciences Würzburg-Schweinfurt
Sanderheinrichsleitenweg 20, 97074 Würzburg, Germany
{tobias.fertig, andreas.schuetz, kristin.weber}@fhws.de

Abstract

Advancing digitization in companies leads to increased significance of information and its security. Since people play a crucial role in protecting information, it is important to sensitize them to information security. Many companies struggle with raising the so-called information security awareness (ISA) in a planned and targeted way. A maturity model (MM) for ISA can help companies to carry out an assessment of the current state regarding ISA and thereby actively manage and plan their future ISA measures. The proposed MM has five maturity levels that were determined mathematically with the help of a polytomous extension of the Rasch model and a hierarchical cluster analysis. Required data for the calculations has been gathered through a survey among 105 organizations. The evaluation has shown that the MM is well-suited to identify strengths and weaknesses with regard to ISA within organizations.

Keywords: Maturity Model, Information Security Awareness, Design Science Research, Rasch Model, Assessment

1. Introduction

The latest report about the state of cybersecurity by ISACA (2022) shows that social engineering (SE) is again the most executed attack. Since SE is only attacking the human factor, information security awareness (ISA) of an organization is very important to prevent those attacks. However, the questions asked by ISACA (2022) about the state of ISA reveal a common lack in organizations: They ask towards the confidence and the feeling about ISA in the participants' organizations - not about the actual state. One reason could be that metrics and measurements in the field of ISA are missing (Fertig & Schütz, 2020).

Besides metrics, one possibility to transform vague feelings into certainty could be an assessment of ISA via standardized guidelines. A Maturity Model (MM) is a tool that may help organizations to assess the current status of their ISA measures. However, according to Fertig et al. (2020) existing information security MM do not consider ISA.

Two of the biggest obstacles according to ISACA (2021) and ISACA (2022) are that an organization requires the expertise to understand and assess maturity, and to assess the maturity of organizational structures, processes, and measures for ISA more often than once per year. In both cases a MM can help: First, organizations can ensure that their measures are targeted and according to appropriate maturity goals. Second, organizations would have a guideline on how to even further advance the maturity of ISA.

This paper presents a MM for ISA which was developed using the research paradigm Design Science (Hevner et al., 2004). We completed the first iteration of a Design Science cycle. First, we conducted systematic literature reviews as well as case studies to analyze requirements of theory and practice for an ISA MM. The MM was evaluated via a focus group and a user workshop. In this paper, we present the first draft of the MM for ISA. We discuss the results of the evaluation of that first draft and how to use them as input for the next Design Science iteration.

After this introduction, we summarize related work and explain the required technical background. In Section 3 we explain the methodology used to answer the research questions. Afterwards, we summarize our conducted study in Section 4 as well as the approach for the development of the MM in Section 5. The

URI: https://hdl.handle.net/10125/103461
978-0-9981331-6-4

last sections will cover the results of the approach, the developed MM, the evaluation, and the discussion. We end with a short conclusion and outlook for future work.

2. Theoretical Background

Information Security Awareness – regarding security threats – is the conscious handling of information, independent of the medium. In order to consider all aspects that influence the conscious handling or the behavior, Schütz (2018) derived the Integrated Behavioral Model (IBM) of Montaño and Kasprzyk (2008) for ISA. The IBM includes different psychological aspects like knowledge, habits, intention, and environmental constraints which all influence compliant behavior.

Many MM for different fields have been proposed. However, many of them share common properties like dimensions, levels and descriptions. According to Fraser et al. (2002) a MM typically has three to six levels, a level descriptor, and a number of dimensions including various elements. Those MM can help assess the abilities of employees as well as organizations (Hakes, 1996; Paulk et al., 1993). Moreover, the continuous improvement of an organization can be evaluated and tracked. However, the systematic literature review by Fertig et al. (2020) showed a lack of sufficient MM for ISA.

In general, MM have been criticized in literature (Mettler, 2009; Pfeffer & Sutton, 1999). According to Biberoglu and Haddad (2002) the most frequent point of criticism is the poor theoretical basis. Another issue is that maturity in information technology is not static and can change over time (Mettler & Pinto, 2018). One example for such a change is shown by ISACA (2021) where 23% of threat actors took advantage of the COVID-19 pandemic.

Additionally, 73% of assessed organizations were stuck in the first level of Capability MM, because it was to hard to meet the requirements of higher levels (Hayes & Zubrow, 1995). To overcome this challenge, Lahrmann et al. (2011) created an approach based on the Rasch model to develop MMs.

Rasch (1960) proposed a psychometric model which aims to analyze categorical data, for example questionnaire responses and answers to questions on a reading assessment. It considers the respondent's abilities and the item difficulty (Rasch, 1960). The item response theory (Hambleton et al., 1991) is the mathematical theory underlying the Rasch model. However, both models have important differences (Linacre, 2005). A central difference lies in the role of specific objectivity which according to Rasch (1977) is a requirement for successful measurement. Since a MM should provide an objective assessment we chose the Rasch model to separate item difficulties and person abilities.

Both Andersen (1977) and Andrich (1978) derived the Rasch model for polytomous response data. With these polytomous extensions the Rasch model can be used with Likert scales. Since the use of a Likert scale (Likert, 1932) provides more context than a dichotomous survey, this extension is especially useful for MMs.

3. Methodology

Within this paper we will answer the following research questions:

RQ1) How can a MM for ISA be developed?

RQ2) How to create a MM which can be used by organizations without getting stuck in their level?

RQ3) How can changes in maturity for information security (IS) be addressed?

The MM was developed via Design Science research which is useful for improving the MM continuously. To create the theoretical basis, we conducted literature reviews and three case studies to gather requirements from theory and practice. The requirements led to an initial draft of a MM. The items of the MM focus for example on the current sensitization measures, the metrics used in the organization, as well as continuous process improvements. Moreover, we clustered those items into four different dimensions: Organization and Management (O), Processes (P), Sensitization (S) and Measuring (M). O focuses on all organizational structures and management behavior regarding IS. P contains procedures and processes which are required for IS within an organization. S assesses how an organization is training its employees, whereas M focuses on the evaluation of success and efficiency of the measures.

Based on the initial draft, we created a survey with LimeSurvey (https://www.limesurvey.org/). For each item we asked two questions: one for the current state of that item within the participant's organization and one for the desired state. The answer possibilities were defined by a 5-step Likert scale (Likert, 1932).

For the data analysis we used a polytomous extension of the Rasch model to support Likert scales (Andrich, 1978). In order to include the current state as well as the desired state, we used the proposed approach of Lahrmann et al. (2011) to derive the item difficulty

based on both values. After calculating the required data via the Rasch model in Winsteps (https://www.winsteps.com/index.htm), we clustered the resulting logits with a hierarchical cluster analysis in SPSS (https://www.ibm.com/products/spss-statistics) via the ward method (Ward, 1963).

This first - mathematically calculated - draft was then evaluated with practitioners. Therefore, we conducted a focus group according to Stewart and Shamdasani (1990) and evaluated the usability of the MM within a workshop on an ISA conference. The gathered feedback and conclusions will be used in the next iteration of the Design Science research.

4. Conducting the Survey

The required survey was carried out using the online survey application LimeSurvey. We ensured an anonymous participation as no personal data and no identifying data regarding the participants' organization were collected. Questions that were not absolutely necessary for the development of the MM have been avoided to reduce the required time for the survey.

The survey contained five pages: One welcome page with some explanations, and four question pages divided into the four MM dimensions (organization, processes, sensitization, and measuring). Each question regarding the current and the desired state had to be answered via a Likert scale with five values (*strongly agree ... strongly disagree*). The Likert scale enables the participants to decide how meaningful they consider each item to be in their organization.

In order to prevent incomplete submissions all questions were declared as mandatory. Aborted surveys were not taken into account during data evaluation. Our aim was to only include complete views of each surveyed organization. Only complete views can be used to achieve a general validity for the MM.

Before we published the survey, we conducted a few test runs. The feedback from the test runs was used to improve our welcome page, all explanations, and the abbreviations, as well as to remove all typos from the questions. Afterwards, the survey was distributed by emailing the survey link to members of various organizations identified as relevant. The target group consists primarily of employees in responsible positions or those with knowledge of the prevailing procedural workflows in the organization (team leaders, department heads, etc.). Moreover, we used newsletter lists to increase the amount of recipients. The survey was distributed to approximately 1,000 people, of which 200 started the survey. We gathered a total of 105 complete submissions. Therefore, the completion rate was round about 10%. Overall, the response rate (i.e. including only partially answered surveys) was about 20%.

We conducted the survey over a period of three months. The first invitations were sent on November 15th, 2021. On February 1st, 2022 we already had 102 complete submissions. Two weeks later only four additional participants started a submission, and three of them completed it. We therefore decided to start the data evaluation on February 15th, 2022. Until the date of paper submission we did not get any additional engagement.

5. Applying the Rasch Model

According to Lahrmann et al. (2011), the following adjustments to the Rasch model are necessary in order to develop a MM. However, these adjustments do not contradict the basic assumptions of the Rasch model (Lahrmann et al., 2011):

The current state A_{vi} of one item i regarding ISA in an organization v would not be sufficient for the creation of a MM. This is due to the fact that ISA is still an emerging topic for organizations. The Rasch model would eliminate items with increased difficulty and would therefore lead to an elimination of currently rarely used ISA principles. For that reason we also considered the desired state D_{vi} for an item i in an organization v. According to the principle of economic efficiency (Samuelson, 1983), the total utility function of an item i does not necessarily increase monotonically in an organization v (Lahrmann et al., 2011). However, D_{vi} could represent an upper limit for this item's increase in utility. The delta X_{vi} from the desired and the actual values then provides desired improvement: $X_{vi} = D_{vi} - A_{vi}$.

Lahrmann et al. (2011) suggested the use of a modified delta. This is due to the fact that MM are supposed to assess the current state, but at the same time should also offer a general and uniform development perspective. This modified delta X_{vi} is the difference between the generally desired value for an item i, $\tilde{D}_i$ (represented by the median value across all organizations) and the respective individual actual values A_{vi} per item and organization: $X_{vi} = \tilde{D}_i - A_{vi}$. The new delta value $\tilde{D}_i$ allows to draw conclusions about the difficulty of an item for the respective organization. A large positive difference between $\tilde{D}_i$ and A_{vi} is a difficult and desirable item, while a negative difference and a difference of 0 are associated with easy items. Since all negative deltas mean over-fulfillment, no further differentiation is made here: $\tilde{D}_i = -2$ as well as $\tilde{D}_i = -1$ are not differentiated and assigned to the range $\tilde{D}_i < 0$. The

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
9	142	105	1.70	.16	1.29	1.35	1.28	1.00	.27	.40	62.9	71.3	O8
35	201	105	.67	.11	1.48	3.04	1.59	2.93	.47	.54	40.0	43.5	M8
25	204	105	.63	.11	.78	-1.67	.73	-1.71	.59	.54	53.3	42.4	S5
20	216	105	.48	.11	.74	-2.15	.74	-1.70	.61	.55	47.6	40.5	P7
24	216	105	.48	.11	1.20	1.45	1.07	.50	.56	.55	45.7	40.5	S4
15	222	105	.42	.11	.64	-3.17	.64	-2.61	.65	.56	56.2	39.9	P8
26	226	105	.37	.11	1.20	1.48	1.47	2.72	.42	.56	37.1	38.8	S6
19	227	105	.36	.11	1.12	.97	1.07	.48	.59	.56	40.0	38.1	P9
7	237	105	.25	.10	.77	-1.93	1.00	.04	.53	.57	41.9	37.9	O7
22	242	105	.19	.10	.87	-1.03	.83	-1.23	.63	.58	41.9	38.0	S2
32	242	105	.19	.10	1.32	2.37	1.39	2.45	.56	.58	35.2	38.0	M4b
33	242	105	.19	.10	1.33	2.41	1.30	1.96	.57	.58	35.2	38.0	M4c
34	244	105	.17	.10	1.38	2.74	1.39	2.45	.51	.58	33.3	38.1	M4d
29	245	105	.16	.10	.97	-.19	.93	-.47	.71	.58	42.9	38.3	M1b
6	249	105	.12	.10	.77	-1.94	.73	-2.02	.66	.58	36.2	38.2	O5
17	251	105	.10	.10	1.04	.35	1.16	1.13	.52	.58	35.2	37.8	P5
5	254	105	.07	.10	.66	-3.13	.63	-2.95	.68	.58	45.7	37.9	O4
23	260	105	.01	.10	.97	-.23	.92	-.56	.62	.59	40.0	37.6	S3
31	260	105	.01	.10	1.44	3.18	1.41	2.68	.60	.59	22.9	37.6	M4a
12	268	105	-.08	.10	.75	-2.22	.74	-2.08	.63	.59	42.9	37.6	P2
2	271	105	-.11	.10	.57	-4.20	.56	-3.87	.70	.59	44.8	36.9	O2
27	276	105	-.16	.10	1.02	.18	1.15	1.11	.57	.59	41.9	37.0	S7
30	279	105	-.19	.10	1.81	5.41	1.77	4.71	.58	.59	22.9	37.0	M2
18	280	105	-.20	.10	.91	-.73	.91	-.67	.63	.60	42.9	36.6	P6
16	284	105	-.24	.10	.76	-2.11	.82	-1.35	.65	.60	39.0	36.7	P4
14	288	105	-.28	.10	.66	-3.08	.64	-3.07	.67	.60	41.0	36.9	P3b
8	297	105	-.37	.10	.65	-3.22	.71	-2.40	.62	.60	51.4	37.4	O10
4	298	105	-.38	.10	.81	-1.64	.77	-1.85	.62	.60	36.2	37.4	O3b
1	307	105	-.47	.10	.68	-2.90	.64	-3.09	.68	.60	47.6	38.0	O1
11	319	105	-.59	.10	.70	-2.63	.69	-2.57	.59	.60	53.3	38.6	P1
10	321	105	-.61	.10	1.10	.80	1.22	1.60	.44	.60	31.4	38.6	O9
28	326	105	-.67	.10	1.61	4.09	1.58	3.71	.57	.60	27.6	39.2	M1a
13	328	105	-.69	.10	.67	-2.93	.70	-2.47	.58	.60	50.5	39.2	P3a
21	328	105	-.69	.10	1.82	5.21	1.74	4.58	.57	.60	21.9	39.2	S1
3	344	105	-.86	.10	.84	-1.25	.80	-1.50	.52	.60	51.4	40.4	O3a
MEAN	262.7	105.0	.00	.10	1.01	-.21	1.02	-.12			41.1	39.4	
P.SD	43.3	.0	.49	.01	.34	2.56	.35	2.36			9.3	5.7	

Figure 1. Item Measure Order for all Items Sorted in Descending Logit

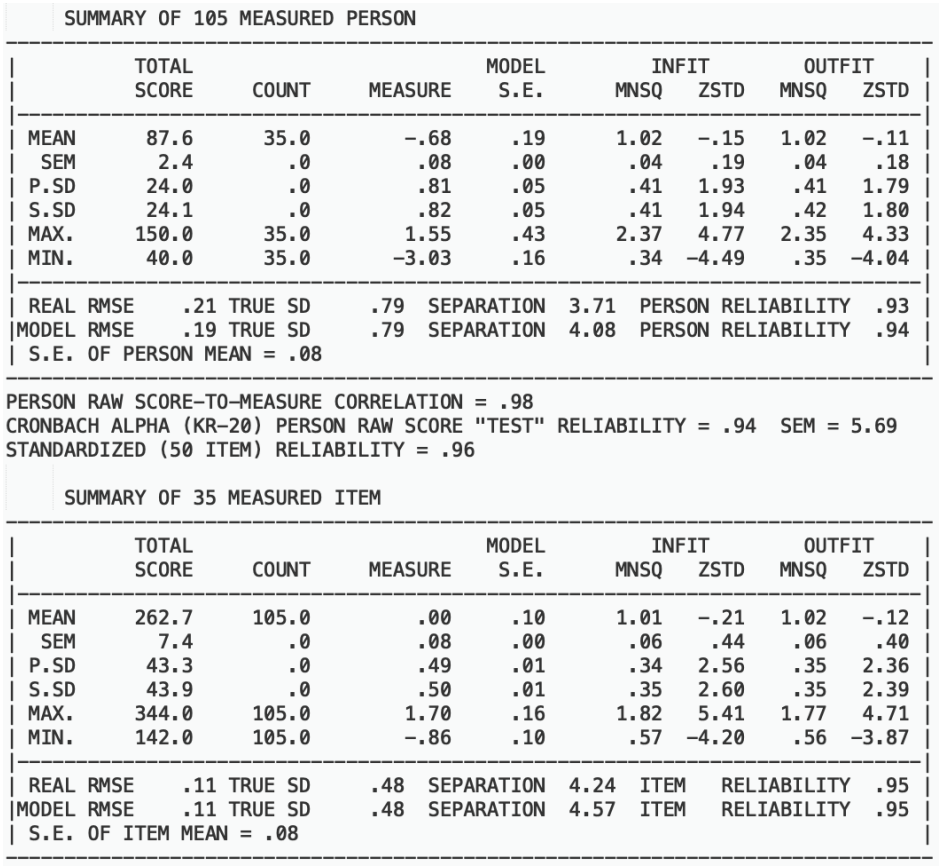
same applies to very large positive delta values: All delta values $D_i > 2$ are aggregated into one category. In order to use the newly defined categories for the Rasch model we transformed the categories into the values 1 to 5 similar to a Likert scale. For a more detailed explanation see Lahrmann et al. (2011).

The Rasch model will result in a list ordered by item difficulty. Since this list alone does not allow a distinct separation into maturity levels, we conducted a clustering. As the MM should have five maturity levels, we defined five as the desired number of clusters. The clustering will determine which intervals of logit values (item difficulty) classify the individual maturity levels from 1 to 5. The cluster analysis was carried out using SPSS (https://www.ibm.com/products/spss-statistics). We used hierarchical clustering with the Ward method (Ward, 1963). Since the delimitation of the maturity levels should be based on the item difficulties, the logit was used as a criterion for clustering. As the clustering was only based on one variable, it was not necessary to carry out any standardization when transforming values.

6. Results of the Rasch Model

Figure 1 shows the measure of item difficulty (JMLE Measure), the standard error (Model SE) and infit and outfit as item statistics. All item statistics were determined with Winsteps (https://www.winsteps.com/index.htm). The item difficulty is called logit. Since higher logit values represent a higher difficulty of the respective item, item O8 is the most difficult item with a logit of 1.70, while item O3a has a logit of -0.86 and represents the easiest of all items. According to Figure 1, the mean value of the logit is 0.00 rounded to two decimal places and thus corresponds exactly to average item difficulty.

Figure 2 shows the separation values. The reliability of the individuals is 0.93 and that of the items is 0.95. Both reliability values represent a high probability that estimated scores will represent the actual scores of a real-life person under assessment. Moreover, high reliability values also indicate that the range of the person measures and number of items were sufficient (Linacre, n.d.-b). The person separation has the value 3.71. Person separations above 2 indicate that an



```
     SUMMARY OF 105 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      87.6      35.0        -.68     .19      1.02   -.15   1.02   -.11 |
|  SEM       2.4        .0         .08     .00       .04    .19    .04    .18 |
| P.SD      24.0        .0         .81     .05       .41   1.93    .41   1.79 |
| S.SD      24.1        .0         .82     .05       .41   1.94    .42   1.80 |
| MAX.     150.0      35.0        1.55     .43      2.37   4.77   2.35   4.33 |
| MIN.      40.0      35.0       -3.03     .16       .34  -4.49    .35  -4.04 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .21 TRUE SD     .79  SEPARATION  3.71  PERSON RELIABILITY  .93 |
|MODEL RMSE    .19 TRUE SD     .79  SEPARATION  4.08  PERSON RELIABILITY  .94 |
| S.E. OF PERSON MEAN = .08                                                   |
-------------------------------------------------------------------------------
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .94  SEM = 5.69
STANDARDIZED (50 ITEM) RELIABILITY = .96

     SUMMARY OF 35 MEASURED ITEM
-------------------------------------------------------------------------------
|          TOTAL                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     262.7     105.0         .00     .10      1.01   -.21   1.02   -.12 |
|  SEM       7.4        .0         .08     .00       .06    .44    .06    .40 |
| P.SD      43.3        .0         .49     .01       .34   2.56    .35   2.36 |
| S.SD      43.9        .0         .50     .01       .35   2.60    .35   2.39 |
| MAX.     344.0     105.0        1.70     .16      1.82   5.41   1.77   4.71 |
| MIN.     142.0     105.0        -.86     .10       .57  -4.20    .56  -3.87 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .11 TRUE SD     .48  SEPARATION  4.24  ITEM   RELIABILITY  .95 |
|MODEL RMSE    .11 TRUE SD     .48  SEPARATION  4.57  ITEM   RELIABILITY  .95 |
| S.E. OF ITEM MEAN = .08                                                     |
-------------------------------------------------------------------------------
```

Figure 2. Separations for Measured Persons and Measured Items

adequate distinction could be made between high and low abilities. The item separation has a value of 4.24. High separation values for items imply that the sample size was big enough to confirm the hierarchy of item difficulties and thus the validity of the construct (Linacre, n.d.-b). All determined values are therefore sufficient.

Figure 1 also shows the mean squared infit and outfit for all items. The mean square values are used to determine how much distortion the measuring system has (Linacre, 2002). The expected value is 1.0 for both infit and outfit. Values < 1.0 indicate that the observations are overly predictable, implying redundancy or model overfit. On the other hand, values > 1.0 indicate unpredictability, suggesting unmodeled noise or a model underfit of the data. The mean squares usually achieve an average value of 1.0 (Linacre, n.d.-a). This is true for the present data as the mean of the infit is 1.01 and the mean of the outfit is 1.02. According to Linacre (1994) the values for infit and outfit in a range from 0.6 to 1.4 are considered appropriate for rating scales in surveys and the range from 0.5 to 1.5 is considered as productive for the measurement. In general, values < 0.5 are considered less productive for the measurement, but have no degrading effect. However, values > 2.0 lead to a distortion or degration of the measurement system (Linacre, 2002).

Nevertheless, neither the infit nor the outfit of any item has a value of < 0.5 or > 2.0. 29 of the 35 item infits are in the appropriate interval range from 0.6 to 1.4. Only three item infits are in the unproductive interval $[1.5; 2.0]$, A similar picture emerges for the outfit: 28 of the 35 item outfits fall within the productive range $[0.6; 1.4]$. Four outfits are in the unproductive interval $[1.5; 2.0]$

Based on the above findings on the fit statistics, no item was identified to be excessively unfit. Therefore, no item was excluded from further use in the MM.

The Wright Map shown in Figure 3 compares the distribution of the measured ability of the respondents on the left side with the distribution of the items on the right side. Since both the item difficulty and the ability of each respondent are expressed in logits on the same scale, they can be compared directly using the Wright Map. The person ability on the left is highest at the top of the scale and lowest at the bottom, with each

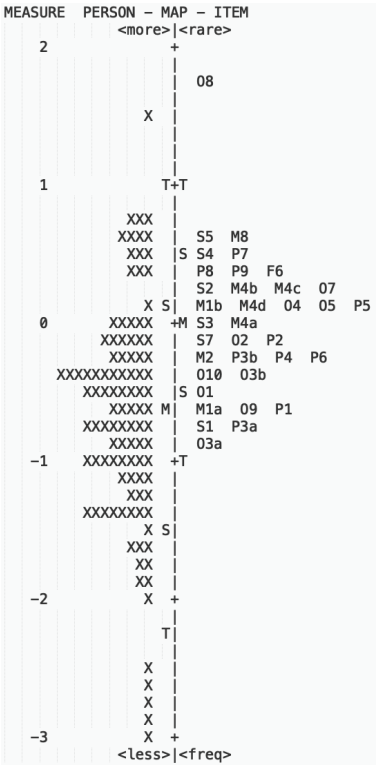


Figure 3. Item Wright Map

"X" representing a person. The items on the right are arranged so that the most difficult items are at the top and the easiest items are at the bottom (Lunz, 2010). All but one item are quite compact in terms of their difficulty in a range from -0.86 to $+0.67$ logit, while only the most difficult item with a logit of 1.70 is significantly higher. In contrast, the skills of the people are distributed over a much larger value range, they extend from -3.03 to $+1.55$ logits.

7. Evaluation

As shown in Section 6, every included item was desired by the participants. Also, the statistics did not require to eliminate any of our proposed items. Therefore, we defined the MM according to the results in Section 6.

Figure 4 shows the resulting MM for ISA after the cluster analysis was conducted. Due to the big gap between the most difficult and the second difficult logit, there is only one item on the fifth maturity level. Moreover, some dimension have no items on some maturity levels. This is due to the nature of the used approach. IDs of items are starting with the letter of their dimension. The numbers in the ID were chosen arbitrarily and have no special meaning.

We scanned the resulting MM for unexpected item placements. The items `P6`, `S7` and `O9` seemed too low in difficulty since all of them focus on continuous improvement processes (CIP). Most MM in literature are placing CIPs on higher levels. Nevertheless, the Rasch model led to the item placement - so we did not change the item rankings. There was no item on a higher level which could be seen as a requirement for another item on a lower level.

After the plausibility check, we conducted a focus group according to the approach of Stewart and Shamdasani (1990). The research agenda for the focus group resulted in the following questions:

G1) Should any item be eliminated from the MM?

G2) Is any item ranked on an inappropriate level?

G3) Are there any items missing?

G4) Are there any dimensions missing?

Afterwards, we prepared the interview guide and recruited ISA experts. A total of 20 experts from industry and academia were invited. In the end, we could recruit six experts for the focus group which is the lower bound according to Stewart and Shamdasani (1990). All experts are currently living in Germany. The focus group lasted for exactly two hours via Zoom (https://zoom.us/). One of us did the moderating task and the other one the recording in writing. The discussion was very intense and led to constructive feedback.

Regarding G1: no item was recommended for elimination. However, some item renaming occurred: First, management should not only know about risks of attacks but also consider them in risk management (Item `O10`). Second, the reward system should be renamed to appreciative behavior (Item `O8`). Third, the workforce should support each other with IS-compliance and not establish a corporate culture where employees observe themselves (Item `P5`). Lastly, according to the experts it does not matter if metrics are measured automatically or manually (Items `M1b, M4a ... M4d`).

Regarding G2: many items were moved around. However, only the items `P6`, `S7` and `O9` were considered far too low and should be placed on Level 4 according to the experts. These were the same items we also would recommend moving up.

Regarding G3: some additional items could be included in the future. The experts recommended additional items for guidelines, corporate strategy, and corporate culture. Moreover, they recommended red teaming as additional item for measuring. They also

Level 1: Initiated		Level 2: Defined		Level 3: Controlled		Level 4: Continuously Improved		Level 5: Appreciated	
Item	Description	Item	Description	Item	Description	Item	Description	Item	Description
Dimension Organization and Management									
O3a	Responsible contact point for ITS concerns of the workforce is available and known	O1	Management considers ISA to be a relevant and important topic	O4	Corporate strategy for ISA is in place			O8	Reward system for the workforce regarding IS-compliant behavior is in place
O9	Employee feedback is used to improve organizational structures	O2	Management is willing to allocate resources to increase ISA	O5	Management communicates the positive contribution of the workforce to the IS of the organization				
		O3b	Responsible contact point for ISA concerns of the workforce is available and known	O7	Management and supervisors exemplify desired IS-compliant behavior for the workforce				
		O10	Management knows the risks of attacks						
Dimension Processes									
P1	Guidelines for IS-compliant behavior of the workforce are defined	P2	The workforce is informed where the defined guidelines can be found	P5	The workforce pays attention to IS-compliant behavior among themselves	P8	The workforce knows which IS guidelines must be observed for new tasks		
P3a	Responsible departments for concerns regarding ITS are communicated to the workforce	P3b	Responsible departments for concerns regarding IS-compliant behavior are communicated to the workforce			P9	Before implementing new activity, the workforce carries out a risk classification regarding IS		
		P4	Defined processes to deal with IS incidents or suspicious cases are in place			P7	Employee feedback is used to improve processes and procedures regarding IS		
		P6	Processes and procedures are analyzed to determine whether they prevent the workforce from complying with IS-compliant behavior						
Dimension Sensitization									
S1	All employees receive at least uniform ISA training	S7	Employee feedback is used to improve targeted sensitization measures	S2	Employees are regularly trained about IS-relevant topics in order to increase the ISA	S4	There are targetted IS trainings that are tailored to the specific needs of the departments in the organization		
				S3	IS-compliant behavior is made prominent and promoted	S5	The workforce is introduced to routines that can lead to IS-compliant habits		
						S6	The workforce receives targetted measures to positively influence the intention of the workforce with regard to IS		
Dimension Measuring									
M1a	Metrics for ITS exist	M2	Sporadic measurements of workforce knowledge about IS are conducted	M1b	Metrics for the automated measurement of IS-compliant behavior exist	M8	Further, manual surveys are conducted based on the findings from metrics		
				M4a	Metrics to automatically measure the knowledge of the workforce regarding IS exist				
				M4b	Metrics for the automated measurement of IS-compliant habits of the workforce exist				
				M4c	Metrics for the automated measurement of the salience of the workforce with regard to IS exist				
				M4d	Metrics to automatically measure the intention of the workforce regarding IS exist				

Figure 4. The Maturity Model for Information Security Awareness (Abbreviations: Information Security (IS), Information Security Awareness (ISA), and IT Security (ITS))

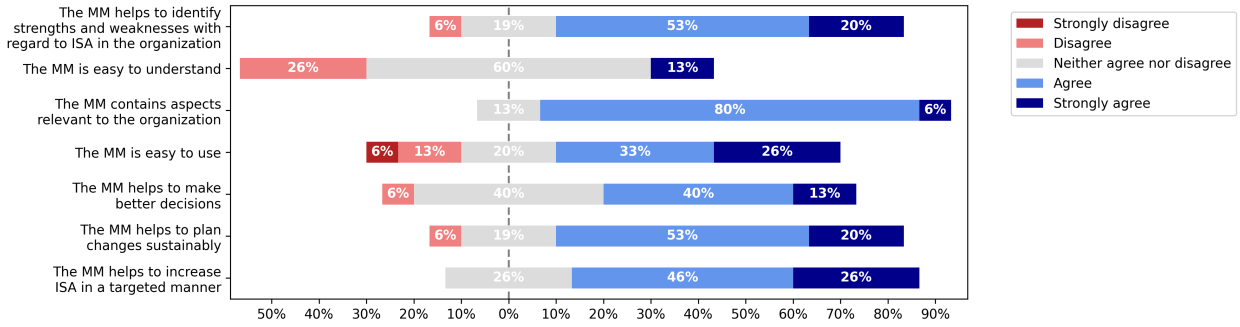


Figure 5. Analysis of the Workshop Survey

recommended to include awards into the sensitization dimension, and to consider an item for agile security.

Regarding G4: no new dimensions were defined. The experts had a discussion about a separate dimension regarding corporate culture. However, they added culture items to organization as well as processes. The experts also discussed to eliminate the dimension measuring since many organizations simply do not measure at all.

After the focus group we wanted to evaluate the MM with potential users. We therefore held a workshop at an ISA conference in Germany[1]. The workshop had 16 participants from different organizations and lasted for about three hours. First, we explained the basics and the methodology used. Afterwards, the participants interviewed each other based on questions for each MM item. After the interviews, we revealed the MM and ranked all organizations according to the interviews. The participants received a ranking of their organization.

At the end we conducted a little survey about the usability of the MM. The survey covered the questions shown in Figure 5. Overall the average response scores are between 3.0 and 4.0. Only the understandability of the MM was rated with an average of 3.0. All other aspects had at least an average rating of 3.6. Most participants agreed that the MM helps to increase IS, to plan changes sustainably and to identify strengths and weaknesses.

8. Discussion and Implications

Within this paper, we explained how we developed a MM for ISA. We therefore defined that each dimension should have 5 levels since this is within the typical range for most MM (see Section 2) and matches the 5-step Likert scale. During the literature research and case studies we gathered different items for the MM. After a clustering of the items we derived the four dimensions shown in Figure 4. Using the dimensions, the items, and the required levels, we were able to develop a MM with the help of the Rasch model and the modifications recommended by Lahrmann et al. (2011).

The experts and the statistical analysis did not eliminate any items. However, the recommended renaming of some items has to be considered in the next iteration of our design science cycle. Some of the items could be moved to higher levels. Nevertheless, we have to consider the movements carefully. We do not want to have organizations getting stuck on low levels. The approach of Lahrmann et al. (2011) considers the current state of organizations for determining the difficulty. Therefore, moving items down without conducting a new survey is problematic. However, we will use this feedback as input for the next Design Science cycle. Within the next iteration we will also include the feedback on missing items. We do not recommend adding new items without conducting another survey. However, if necessary, including new items on the highest level would not prevent an organization from advancing. Since the experts agreed on distributing culture items between the dimensions organization and processes, we will definitely research this topic further within the next iteration. We did not eliminate the dimension regarding measuring, since a MM should also support maturing and improving the current state.

We reviewed the rankings of the participants' organizations after the user workshop. Some of the organizations have already implemented most of the items from Level 1 to 4. However, around 50% were stuck on Level 1 because of a few items. We checked which items were preventing the organizations and discovered in all those cases again the same previously mentioned items `P6`, `S7`, and `O9`. This indicates that the experts' recommendation to move those items to a higher level seems to be useful. The maturity of the participants' organizations was rather high, and they were are already very advanced in ISA. All of the

[1] https://www.take-aware-events.com/events/take-aware-und-sexy-security-2022

organizations participating in the workshop have been trying to improve their ISA for many years. We assume that organizations in general would not achieve such high maturity rankings.

The use of the Rasch model together with the introduced desired state allowed us to develop a MM for ISA (RQ1). We tried to reduce the amount of organizations getting stuck on low levels (RQ2). According to the participants' ranking, the MM still requires some improvements in this case. However, we could already consider the current skill-level of organizations due to the Rasch model. Of course the skill-levels can change and increase over time. Moreover, the maturity requirements for IS also can change over time (RQ3). This is why we recommend repeating the survey in the future to adapt the MM to the future requirements of organizations. The MM should not be considered as final, since IS is always changing. According to Normann Andersen et al. (2020) "the purpose of a MM is not to provide absolute truth, but possibly to provide a useful instrument to practitioners in comprehending and dealing with the difficult task of digital transformation. In this sense, the development and evaluation of maturity models should not necessarily follow a positivistic approach, but rather a design-oriented paradigm." This is why we will improve the MM in the next iteration of our Design Science research. Adapting the MM in iterations will also allow us to consider latest topics, increases in skill-levels, and maintain the usefulness of the MM. We will use the feedback and outcomes as input for the next iteration.

To sum up, the conducted research leads to the following implications for practice: The MM provides a structured guideline on how to advance or improve ISA within organizations. According to the participants of our workshop, the MM contains relevant aspects for their organizations. As shown in Figure 5 most of the participants think the MM is already helping in its current state.

Moreover, the MM allows to assess ISA and therefore, the human factor. This is an advantage to existing MM for IS which are ignoring ISA (Fertig et al., 2020). Since our MM is solely focusing on ISA, it can easily be used in addition to already existing MM for technological aspects of IS.

ISACA (2021) concluded that the importance of assessing IS maturity is unquestioned. The human factor is an important part of IS, so the assessment of IS maturity should also include ISA. Since many organizations are already assessing their IS annually (ISACA, 2022) the MM can provide a tool to complement ISA within their assessments. An easy-to-use MM would reduce the required efforts. Moreover, the organizations can increase their IS in a targeted manner and track the progress via consistent assessments. Those assessments could then be used to conduct benchmarks - either between departments or between multiple organizations. In the end, reports like the one from ISACA (2022) would no longer have to rely on confidence and feelings.

9. Conclusion

In order to develop a MM for ISA, we decided to use Design Science as research paradigm. Design Science supports the continuous development of an artifact through multiple iterations. With that, we can improve the MM with each iteration due to the changing maturity requirements (Mettler & Pinto, 2018). We conducted literature reviews as well as case studies to fulfill the rigor and relevance cycle of Design Science.

After the definition of requirements we developed the MM with the help of the approach of Lahrmann et al. (2011). We conducted a survey in organizations to uncover their current and desired state regarding ISA. We had 105 complete responses which were then analyzed using the Rasch model (Rasch, 1960). Since the Rasch model only delivers all items ranked by difficulty, we had to do a cluster analysis in SPSS. Finally, the resulting five clusters defined the classification of the maturity items into the five levels of the MM.

For evaluation purposes we conducted a focus group with experts in the field of ISA. The experts were from industry and academia. We also tested our MM at an ISA workshop with possible users.

During both evaluations we gathered insights on how to improve the MM. The current MM as well as all feedback will serve as input for the next Design Science iteration. The next steps will be to test the MM in some organizations to gather additional requirements. We will also repeat the literature review to uncover more information about corporate culture. The MM will change with each iteration, which allows it to remain relevant for organizations. At the end, the MM for ISA can be a useful tool to improve the overall security within an organization and a useful guideline for planning next steps within an organization.

Acknowledgements

Tobias Fertig and Andreas E. Schütz were supported by the BayWISS Consortium Digitization.

References

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*(1), 69–81. https://doi.org/10.1007/BF02293746

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Biberoglu, E., & Haddad, H. (2002). A survey of industrial experiences with CMM and the teaching of CMM practices. *Journal of Computing Sciences in Colleges*, *18*(2), 143–152.

Fertig, T., & Schütz, A. (2020). About the Measuring of Information Security Awareness: A Systematic Literature Review. *53rd Hawaii International Conference on System Sciences*. Retrieved January 20, 2020, from http://scholarspace.manoa.hawaii.edu/handle/10125/64540

Fertig, T., Schütz, A. E., Weber, K., & Müller, N. H. (2020). Towards an Information Security Awareness Maturity Model. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Human and Technology Ecosystems* (pp. 587–599). Springer International Publishing. https://doi.org/10.1007/978-3-030-50506-6_40

Fraser, P., Moultrie, J., & Gregory, M. (2002). The use of maturity models/grids as a tool in assessing product development capability. *IEEE International Engineering Management Conference*, *1*, 244–249 vol.1. https://doi.org/10.1109/IEMC.2002.1038431

Hakes, C. (1996). *The corporate self assessment handbook: For measuring business excellence*. Springer Netherlands. https://books.google.de/books?id=MKzjAAAACAAJ

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. SAGE.

Hayes, W., & Zubrow, D. (1995). *Moving on up: Data and experience doing cmm-based process improvement* (tech. rep. CMU/SEI-95-TR-008). Software Engineering Institute, Carnegie Mellon University. Pittsburgh, PA. http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=12353

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

ISACA. (2021). *State of Cybersecurity 2021. Part 2: Threat Landscape, Security Operations and Cybersecurity Maturity* (tech. rep.) [Published: ISACA, Golf Road, Suite 400, Schaumburg, IL 60173 USA]. Information Systems Audit and Control Association. https://www.isaca.org/go/state-of-cybersecurity-2021

ISACA. (2022). *State of Cybersecurity 2022* (tech. rep.) [Published: ISACA, Golf Road, Suite 400, Schaumburg, IL 60173 USA]. Information Systems Audit and Control Association. https://www.isaca.org/go/state-of-cybersecurity-2022

Lahrmann, G., Marx, F., Mettler, T., Winter, R., & Wortmann, F. (2011). Inductive Design of Maturity Models: Applying the Rasch Algorithm for Design Science Research. In H. Jain, A. P. Sinha, & P. Vitharana (Eds.), *Service-Oriented Perspectives in Design Science Research* (pp. 176–191). Springer. https://doi.org/10.1007/978-3-642-20633-7_13

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22 140*, 55–55.

Linacre, J. M. (n.d.-a). Diagnosing Misfit [Winman] [Last accessed: 2022-06-08]. https://www.winsteps.com/facetman/diagnosingmisfit.htm

Linacre, J. M. (n.d.-b). Reliability and separation of measures [Winman] [Last accessed: 2022-06-08]. https://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (1994). Reasonable mean-square fit values. [Rasch measurement transactions] [Last accessed: 2022-06-08]. https://www.rasch.org/rmt/rmt83b.htm

Linacre, J. M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? [Rasch measurement transactions] [Last accessed: 2022-06-08]. https://www.rasch.org/rmt/rmt162f.htm

Linacre, J. M. (2005). Rasch Dichotomous Model vs. One-parameter Logistic Model (1PL 1-PL) [Rasch measurement transactions] [Last accessed: 2022-06-08]. https://www.rasch.org/rmt/rmt193h.htm

Lunz, M. E. (2010). Measurement Research Associates Test Insights [Last accessed: 2022-06-08]. https://www.rasch.org/mra/mra-01-10.htm

Mettler, T. (2009). A Design Science Research Perspective on Maturity Models in Information Systems. Retrieved June 10, 2022, from https://www.alexandria.unisg.ch/214531/

Mettler, T., & Pinto, R. (2018). Evolutionary paths and influencing factors towards digital maturity: An analysis of the status quo in Swiss hospitals. *Technological Forecasting and Social Change*, *133*, 104–117. https://doi.org/10.1016/j.techfore.2018.03.009

Montaño, D. E., & Kasprzyk, D. (2008). Theory of Reasoned Action, Theory of Planned Behavior, and the Integrated Behavior Model. In K. Glanz, Rimer, Barbara, K., & K. Viswanath (Eds.), *Health Behavior and Health Education* (pp. 67–96). APA PsycNet.

Normann Andersen, K., Lee, J., Mettler, T., & Moon, M. J. (2020). Ten Misunderstandings about Maturity Models. *The 21st Annual International Conference on Digital Government Research*, 261–266. https://doi.org/10.1145/3396956.3396980

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. (1993). *Capability Maturity Model for Software (Version 1.1)* (tech. rep. CMU/SEI-93-TR-024). Carnegie Mellon University. Retrieved January 20, 2020, from https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=11955

Pfeffer, J., & Sutton, R. I. (1999). Knowing "What" to do is not Enough: Turning Knowledge into Action. *California Management Review*, *42*(1), 83–108. https://doi.org/10.1177/000812569904200101

Rasch, G. (1977). On Specific Objectivity: An Attempt of Formalizing the Generality and Validity of Scientific Statements. *Danish Yearbook of Philosophy*, *14*, 58–94.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Samuelson, P. A. (1983). *Foundations of Economic Analysis*. Harvard University Press.

Schütz, A. E. (2018). Information Security Awareness: It's Time to Change Minds! *Proceedings of International Conference on Applied Informatics Imagination, Creativity, Design, Development - ICDD 2018*.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice*. Sage Publications, Inc.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845