## APPENDIX A

The dataset of professional recipes and the dataset of amateur recipes each have their own three-layer tree structure of sub-categories. For both dataset, one recipe originally corresponds to one sub-category (e.g., tomatoes, mushroom, chocolate cakes).

In this study, each experiment was designed so that the work time per participant would be limited to approximately one hour taking into consideration the physical and mental burden on the participants.

## APPENDIX A.1 COLLECTION

In this *Collection* step, the two recipes presented each task for each participant were randomly selected from those corresponded to the same sub-category in the professional and amateur recipe datasets. It is because we assumed that participants would not be able to think deeply about what makes each recipe useful if they compare two very different recipes (e.g., a chocolate cake recipe and a curry recipe). We have to note that we used the professional and amateur recipes were mixed and the two recipes were randomly selected from this mixed dataset; therefore, participants did not necessarily compare the professional recipes with the amateur recipes.

The details of the definitions of the number of participants are as follows:

(20 participants) × (20 tasks (pair of recipes))/(67 pairs of recipes) = 5.97, therefore, approximately five participants, on average, checked each pair of recipes.

## APPENDIX A.2 AGGREGATION

In this *Aggregation* step, the details of the definitions of the number of participants are as follows:

The number of combinations of 426 content candidates are ${}_{426}C_{2} = 90{,}525$.

370 (participants) × 25 (tasks) × 100 (pairs of content candidates)/90,525 = 10.22, therefore, approximately 10 participants, on average, checked each pair of content candidate.

It is because future research might involve supporting people's efforts to describe the extracted content categories, the threshold value of the similarity (i.e., "Pairs of content candidates that four or more study participants judged to be similar were considered similar") was set so that the number of content categories (see Section 3.1.3) would be between five and nine [ref1], which is the number of elements that humans can retain using short-term memory.

[ref1] Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev. 63, 2, 81–97. https://doi.org/10.1037/h0043158

## APPENDIX A.3 CONCEPTUALIZATION

In this *Conceptualization* step, regarding the categories about content, only "reasons for tools used" was renamed to "reasons for the cooking process," based on this study's intent.

## APPENDIX B

**Table A-1. The number of each type of experimental documents created for each content category**

| | | Arrangement | | | | |
|---|---|---|---|---|---|---|
| Content category | Deleted document | A01 | A02 | A03 | A04 | A05 |
| Reasons | 13 | 13 | 9 | 5 | 13 | 13 |
| Suitable ingredients | 40 | 40 | 40 | 40 | 40 | 40 |
| Notes | 40 | 39 | 40 | 32 | 40 | 40 |
| Substitutions | 21 | 21 | 21 | 21 | 21 | 21 |
| Variations | 44 | 44 | 44 | 17 | 44 | 44 |
| Servings | 30 | 30 | 30 | 1 | 30 | 30 |

When the deleted documents were created, a minimal expression adjustment was made only when the document became grammatically incorrect when the sentences indicating the corresponding content category were deleted. When the rearranged documents were created, if a rearranged document was the same as the original, it was not created. We did not create deleted or rearranged documents for content categories that were not included in the relevant original document.

As a result, the number of each type of experimental documents (recipe) created for each content category were shown in the Table A-1. To avoid bias in the number of times each arrangement of content categories was used in the experiments in Section 4.1.2, we randomly selected 13 experimental documents from all the arrangements of content categories in Table 1 and used them in the experiments in Section 4.1.2. For the arrangement of content categories Ressons_A02, Reasons_A03, and Servings_A03, for which there were fewer than 13 experimental documents, all the corresponding experimental documents were used. Therefore, a total of 444 pairs of the deleted document and the original document or five types of rearranged documents were discussed in this analysis.

300 (participants) × 20 (tasks for Step 1 or 2)/444 tasks = 16.89, therefore, approximately over 15 participants, on average, checked each pair of content candidate.

**APPENDIX C**

Raw data (7,500 answers)

| participant_ID | usefulness score | unknown words | understandability | skill level | Information needs (Reasons) | Information needs (Suitable Ingredients) | … |
|---|---|---|---|---|---|---|---|
| Id_001 | 1 | 1 | 8 | 8 | 8 | 7 | |
| Id_001 | 2 | 0 | | 7 | 8 | 7 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| Id_001 | 0 | 1 | 10 | 6 | 8 | 7 | |
| Id_002 | -2 | 0 | 9 | 8 | 9 | 9 | |
| Id_002 | 1 | | 8 | 6 | 9 | 9 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

Each participant repeated 25 tasks, and 300 participants answered.

There were some defects.

Data used for Analysis in Section 4.2 (5,574 answers)

| participant_ID | usefulness score | unknown words | understandability | skill level | Information needs (Reasons) | Information needs (Suitable Ingredients) | … |
|---|---|---|---|---|---|---|---|
| Id_001 | 1 | 1 | 8 | 8 | 8 | 7 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| Id_001 | 0 | 1 | 10 | 6 | 8 | 7 | |
| Id_002 | -2 | 0 | 9 | 8 | 9 | 9 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

**Figure A-1. The details of raw data obtained in Section 4.1. and the data used in Section 4.2.**

**APPENDIX D**

**APPENDIX D.1**

We checked which data among participants' understanding and information needs were hierarchical with respect to the usability score. The interclass correlation coefficient (ICC) and the design effect (DE) [ref2] for each data were as follows. For the presence or absence of unknown words, ICC was 0.03 and DE was 17.3. For the number of unknown words, ICC was 0.08 and DE was 10.1. For the understandings, ICC was 0.00 and DE was 1.00. For the skill levels, ICC was 0.00 and DE was 1.10. For the Information needs for substitutions, ICC was 0.00 and DE was 1.00. For the Information needs for variations, ICC was 0.08 and DE was 10.0. For the Information needs for reasons, ICC was 0.01 and DE was 2.17. For the Information needs for suitable substitutions, ICC was 0.0 and DE was 1.0. For the Information needs for suitable substitutions, ICC was 0.0 and DE was 1.0. For the Information needs for notes, ICC was 0.00 and DE was 1.5. For the Information needs for servings, ICC was 0.01 and DE was 2.2.

From the above, the usefulness scores were judged to show hierarchy among the number of unknown words of participants' mostly among participants' understandings and information needs.

[ref2] Hox, Joop J., Moerbeek, Mirjam, Van de Schoot, Rens. 2017. Multilevel Analysis: Techniques and Applications (3rd. ed.). Routledge, New York, NY.

**APPENDIX D.2 Methods**

We estimated parameters the below models for each content category.

- Model original, Model A01, Model A02, Model A03, Model A04, Model A05

$$usefullness_{ij} = \alpha + \sum_{k=1}^{6} \beta_k n_{ik} + \eta_k^{(z_{ijk})} + \gamma_k^{(z_{ijk})} n_{ik} + e_{ij} \ (k = 1,..,6) \quad (1)$$

$$e_{ij} \sim N(0, \sigma_\gamma^2) \quad (2)$$

For the usefulness score $usefullness_{ij}$ given to the $i$-th document by the $j$-th participants, $\alpha$ was the intercept, $k$ meant the arrangement of each content category (original, A01, A02, A03, A04, A05), $\beta_k$ $(k = 1,..,6)$ was the coefficient for $n_{ki}$, and $n_{ki}$ was the number of descriptions for each of the six content categories. For Model original, Model A01, Model A02, Model A03, Model A04, and Model A05, $k$ meant original document, A01, A02, A03, A04, and A05, respectively. Specifically, we used non-informative prior for $\beta_k$, $\alpha \sim \text{StudentT}(3, 0, 2.5)$, and $\sigma_e \sim \text{StudentT}(3, 0, 2.5)$ as the prior distributions. $z_{ijk} \in \{0, 1\}$ indicated whether the arrangement (e.g., for Model original, original. For Model A01, A01.) were used for the description of the $k$-th arrangement in the $i$-th document for the $j$-th participants and $\eta_k^{(z_{ijk})}$ was the random effect of the arrangement on the intercept for the $k$-th content category of the $i$-th document. $\gamma_k^{(z_{ijk})}$ was the random effect of the arrangement on the coefficient for $n_{ik}$. The models used the same priors as Model base. Non-informative prior was used for the prior distribution of SDs of random effects, and $LKJCholesky(1)$ was the prior of the correlation matrix between $\gamma_k^{(g)}$ and $\eta_k^{(g)}$ for $k \in \{1, \ldots, 6\}$ and $g \in \{0, 1\}$.

The model parameters were fitted with four Markov chain Monte Carlo (MCMC) chains with 10,000 iterations and 5,000 burn-in samples with a thinning parameter of one.

**APPENDIX D.3 Results**

The WAIC scores of each model are shown in Table A-2. All $\hat{R}$ were under 1.05.

**Table A-2.** WAIC values for each model

| | Reasons | Suitable ingredients | Notes | Substitutions | Variations | Servings |
|---|---|---|---|---|---|---|
| Model original | 14744.9 | 14742.1 | **14674.0** | 14742.0 | **14748.5** | 14752.0 |
| Model A01 | 14746.7 | 14745.4 | 14681.3 | 14742.1 | 14749.0 | 14753.0 |
| Model A02 | **14744.2** | **14741.3** | 14687.2 | 14741.3 | 14750.6 | **14751.3** |
| Model A03 | 14746.5 | 14733.3 | 14680.7 | 14741.7 | 14747.1 | 14751.6 |
| Model A04 | 14745.5 | 14741.7 | 14681.3 | 14746.8 | 14749.7 | 14748.6 |
| Model A05 | 14746.6 | 14742.1 | 14682.3 | **14741.2** | 14750.1 | **14751.3** |

The details of the results for the model with the smallest WAIC for each content category were shown in Table A-3 – A-8.

**Table A-3.** The estimation results of Model A02 for Reasons

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.37 | 0.60 | -1.01 | 1.63 | 1.00 |
| original ($\beta_1$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A01 ($\beta_2$) | -0.01 | 0.01 | -0.03 | 0.00 | 1.00 |
| A02 ($\beta_3$) | -0.03 | 0.78 | -1.83 | 1.67 | 1.00 |
| A03 ($\beta_4$) | 0.01 | 0.01 | -0.01 | 0.04 | 1.00 |
| A04 ($\beta_5$) | -0.01 | 0.01 | -0.03 | 0.00 | 1.00 |
| A05 ($\beta_6$) | -0.01 | 0.01 | -0.02 | 0.01 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_3$ ) | 0.63 | 0.99 | 0.01 | 3.43 | 1.01 |
| sd ($\gamma_3$ ) | 0.81 | 1.04 | 0.01 | 3.78 | 1.00 |
| Corr ($\eta_3$ , $\gamma_3$ ) | 0.02 | 0.63 | -0.96 | 0.98 | 1.00 |

**Table A-4.** The estimation results of Model A03 for Suitable ingredients

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.32 | 0.61 | -1.15 | 1.53 | 1.00 |
| original ($\beta_1$) | -0.02 | 0.01 | -0.03 | -0.00 | 1.00 |
| A01 ($\beta_2$) | 0.01 | 0.01 | **0.00** | **0.03** | 1.00 |
| A02 ($\beta_3$) | -0.01 | 0.01 | -0.02 | 0.01 | 1.00 |
| A03 ($\beta_4$) | 0.05 | 1.12 | -2.36 | 2.55 | 1.00 |
| A04 ($\beta_5$) | -0.00 | 0.01 | -0.02 | 0.01 | 1.00 |
| A05 ($\beta_6$) | -0.00 | 0.01 | -0.02 | 0.01 | 1.00 |
| Family Specific Parameters: | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_4$ ) | 0.71 | 1.02 | 0.01 | 3.45 | 1.00 |
| sd ($\gamma_4$ ) | 1.12 | 1.28 | 0.08 | 4.73 | 1.00 |
| Corr ($\eta_4$ , $\gamma_4$ ) | -0.08 | 0.60 | -0.98 | 0.95 | 1.00 |

**Table A-5.** The estimation results of Model original for Notes

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.31 | 0.63 | -1.27 | 1.46 | 1.00 |
| original ($\beta_1$) | -0.05 | 0.81 | -1.96 | 1.71 | 1.00 |
| A01 ($\beta_2$) | 0.03 | 0.01 | **0.01** | **0.04** | 1.00 |
| A02 ($\beta_3$) | -0.04 | 0.01 | -0.05 | -0.02 | 1.00 |
| A03 ($\beta_4$) | 0.03 | 0.01 | **0.01** | **0.04** | 1.00 |
| A04 ($\beta_5$) | 0.02 | 0.01 | **0.01** | **0.03** | 1.00 |
| A05 ($\beta_6$) | 0.01 | 0.01 | -0.01 | 0.02 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.90 | 0.01 | 0.88 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_1$ ) | 0.66 | 1.00 | 0.01 | 3.42 | 1.00 |
| sd ($\gamma_1$ ) | 0.83 | 1.01 | 0.04 | 3.81 | 1.00 |
| Corr ($\eta_1$ , $\gamma_1$ ) | 0.01 | 0.63 | -0.98 | 0.98 | 1.00 |

**Table A-6.** The estimation results of Model A05 for Substitutions

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.32 | 0.71 | -1.24 | 1.79 | 1.00 |
| original ($\beta_1$) | -0.01 | 0.01 | -0.02 | 0.00 | 1.00 |
| A01 ($\beta_2$) | -0.01 | 0.01 | -0.02 | 0.01 | 1.00 |
| A02 ($\beta_3$) | -0.01 | 0.01 | -0.03 | 0.00 | 1.00 |
| A03 ($\beta_4$) | -0.00 | 0.01 | -0.02 | 0.01 | 1.00 |
| A04 ($\beta_5$) | -0.00 | 0.01 | -0.02 | 0.01 | 1.00 |
| A05 ($\beta_6$) | -0.05 | 0.72 | -2.08 | 1.50 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_6$ ) | 0.70 | 1.07 | 0.01 | 3.67 | 1.01 |

| | | | | | |
|---|---|---|---|---|---|
| sd ($\gamma_6$ ) | 0.75 | 1.14 | 0.01 | 4.02 | 1.00 |
| Corr ($\eta_6$ , $\gamma_6$ ) | -0.04 | 0.63 | -0.98 | 0.97 | 1.00 |

**Table A-7.** The estimation results of Model original for Variations

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.37 | 0.60 | -1.03 | 1.65 | 1.00 |
| original ($\beta_1$) | -0.03 | 0.71 | -1.57 | 1.54 | 1.00 |
| A01 ($\beta_2$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A02 ($\beta_3$) | -0.01 | 0.01 | -0.03 | 0.00 | 1.00 |
| A03 ($\beta_4$) | -0.00 | 0.01 | -0.02 | 0.01 | 1.00 |
| A04 ($\beta_5$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A05 ($\beta_6$) | -0.00 | 0.01 | -0.01 | 0.01 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_1$ ) | 0.62 | 0.94 | 0.01 | 3.31 | 1.00 |
| sd ($\gamma_1$ ) | 0.64 | 0.90 | 0.01 | 3.23 | 1.00 |
| Corr ($\eta_1$ , $\gamma_1$ ) | 0.00 | 0.63 | -0.98 | 0.98 | 1.00 |

**Table A-8_1.** The estimation results of Model A02 for Servings

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.38 | 0.53 | -0.82 | 1.58 | 1.00 |
| original ($\beta_1$) | -0.00 | 0.01 | -0.01 | 0.01 | 1.00 |
| A01 ($\beta_2$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A02 ($\beta_3$) | -0.01 | 0.66 | -1.64 | 1.33 | 1.00 |
| A03 ($\beta_4$) | 0.00 | 0.03 | -0.05 | 0.07 | 1.00 |
| A04 ($\beta_5$) | 0.01 | 0.01 | -0.00 | 0.02 | 1.00 |
| A05 ($\beta_6$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_3$ ) | 0.60 | 0.92 | 0.01 | 3.22 | 1.00 |
| sd ($\gamma_3$ ) | 0.64 | 0.97 | 0.01 | 3.42 | 1.00 |
| Corr ($\eta_3$ , $\gamma_3$ ) | -0.04 | 0.62 | -0.98 | 0.97 | 1.00 |

**Table A-8_2.** The estimation results of Model A05 for Servings

| Population-Level Effects: | | | | | |
|---|---|---|---|---|---|
| | Estimate | Est.Error | l-95% CI | U-95% CI | Rhat |
| Intercept | 0.41 | 0.61 | -0.81 | 1.72 | 1.00 |
| original ($\beta_1$) | -0.00 | 0.01 | -0.01 | 0.01 | 1.00 |
| A01 ($\beta_2$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A02 ($\beta_3$) | 0.00 | 0.01 | -0.01 | 0.02 | 1.00 |
| A03 ($\beta_4$) | 0.01 | 0.03 | -0.04 | 0.07 | 1.00 |
| A04 ($\beta_5$) | 0.01 | 0.01 | -0.00 | 0.03 | 1.00 |
| A05 ($\beta_6$) | 0.01 | 0.63 | -1.29 | 1.57 | 1.00 |
| Family Specific Parameters: | | | | | |
| sigma | 0.91 | 0.01 | 0.89 | 0.92 | 1.00 |
| Group-Level Effects | | | | | |
| the presence of unknown words in Substitute | | | | | |
| sd ($\eta_6$ ) | 0.61 | 0.88 | 0.01 | 3.18 | 1.00 |
| sd ($\gamma_6$ ) | 0.66 | 0.92 | 0.01 | 3.35 | 1.01 |
| Corr ($\eta_6$ , $\gamma_6$ ) | 0.02 | 0.63 | -0.97 | 0.98 | 1.00 |