Electronic Supplementary Material (ESI) for Chemistry Education Research and Practice.

Appendix 1: Research Instruments

The learning event consisted of a textbook passage with a general description of the Second Law of Thermodynamics and common to all the participants. A self-explaining task, SE-Task, followed this passage. There were four different SE-Tasks, each defining one of the study conditions: (1) Self-explaining own answer, SEA; (2) Self-explaining agreement/disagreement, EADA; (3) Self-explaining for others, SEO; and (4) Self-explaining incorrect answer, SEIA. Following information provided to students during the learning event.

Second Law of Thermodynamics (General description)

We have seen that both the system and surroundings may undergo changes in entropy during a process. The sum of the entropy changes for the system and the surroundings is the entropy change for the universe:

$$\Delta S_{univ} = \Delta S_{sys} + \Delta S_{surr}$$

The Second Law of Thermodynamics says that for a process to be spontaneous as written (in the forward direction), ΔS_{univ} must be positive ($\Delta S_{univ} > 0$). Therefore, the system may undergo a decrease in entropy as long as the surroundings undergo a larger increase in entropy making the resulting ΔS_{univ} positive, and vice versa. A process for which ΔS_{univ} is negative is not spontaneous as written.

Self-explaining tasks

1. **Self-explaining own answer (SEA): Working on a problem and explaining one's own answer.**

> When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). However, this process is spontaneous. How do you explain this? Please be as thorough in your response as possible.

2. **Self-explaining agreement/disagreement (EADA): Considering others' answers to a problem and explaining one's agreement/disagreement.**

> When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). Despite this observation, your group members maintain that this process is spontaneous. Therefore, they say, no energy input from the outside is necessary to make this change happen. Do you agree with your classmates? Please explain and be as thorough in your response as possible.

3. **Self-explaining for others (SEO): Explaining answer to a problem for others to use in their studying.**

When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). However, this process is <u>spontaneous</u>. Explain this in writing so that a classmate of yours can use your explanation as reference when answering a similar problem. Your answer will be used by your classmate. Please be as thorough in your response as possible.

4. **Self-explaining incorrect answer (SEIA): Explaining others' incorrect answers to a problem.**

When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). Your group members maintain that this process <u>will not occur spontaneously</u>. Therefore, they say, there must be an energy input from the outside to make this change happen; otherwise, water will not freeze. This stance is incorrect. What do you think led your classmates to this incorrect conclusion? Please be as thorough in your response as possible.

Appendix 2: Interview protocol

Think-aloud Interview protocol

Students' assessment of research materials

1. Introductory aspects [read to interviewee]

We are testing an instrument that has questions that may be difficult to understand, hard to answer, or that make little sense. We would like you to answer the questions as carefully as possible. *We are primarily interested in the ways that they arrived at those answers, and the problems they encountered*. Therefore, any detailed help you can give us is of interest, even if it seems irrelevant or trivial.

We are not looking for correct answers; we just want to listen to your comments. I didn't write these questions, so don't worry about hurting my feelings if you criticize them -my job is to find out what's wrong with them.

The conversations will be audio taped just as a means for us to go back and review *what* was said and not *who* said *what*. This interview is confidential; you will not be identified by name and only the transcriber will listen to this tape. The transcriber is bound to confidentiality, as well. During the conversation, I may take notes, which most probably will be reminders to myself of something I want to inquire about later, or something especially interesting you said. I will not jot down things *about* you, you are not under observation.

Please feel free to spend as much time as you need or want on any given topic. You do not have to reply to a question if for any reason you do not feel comfortable. We may stop the conversation at any time you wish or need to. Do not feel like I am being too insistent if I ask some follow up questions to your comments. It is our interest to clearly understand what you mean; we are trying to get to a deeper level of understanding.

Once again, **this interview is absolutely confidential.** We very much appreciate your taking the time for this conversation. We will start with some general background information and then we will move on to aspects related to the instrument.

2. **Background [use these to strengthen rapport with interviewee and set a comfortable environment]**
    a) What is your undergraduate major in?
    b) What chemistry courses have you taken in the past?
    c) Are you taking any chemistry classes this semester?

3. **Think-aloud training exercise:**

"Try to visualize the place where you live, and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about."

4. **Instrument assessment. One self-explaining task condition. [Use prompts and follow-ups as necessary]**
    1. The following instrument is intended for students taking general chemistry 2. Please read the following information.
    2. Give student information sheet "Entropy definition". Give time to read. Then remove information.
    3. Give student information sheet "Second law of thermodynamics. Give time to read. Let student keep this information for the rest of the interview.
    4. I will read a question to you and I would like you to think out loud when you answer the following questions.
    5. Read prompt "Self-explaining own answer, SEA" to student to think-aloud while solving it.

Verbal Probes during resolution:

a. Please repeat the question I just asked in your own words?
b. How did you arrive at that answer?
c. I noticed that you hesitated - tell me what you were thinking.

6. Verbal Probes after resolution:
    a. How difficult was this question to answer?
    b. How sure are you of your answer?

5. **Instrument assessment. Comparison with two other intervention conditions. [Use prompts and follow-ups as necessary]**
   1. Now I am going to give you another question.
   2. Provide prompt “Self-explaining agreement/disagreement, EADA” to student. Give time to read. Let student keep this information for the rest of the interview. Verbal probe technique is used.

Verbal Probes:

a. What does the term "energy input from the outside" mean to you?
b. How hard is it to think of reasons for your classmates to get the incorrect conclusion?
c. What other reasons can you think of?
d. Overall, how difficult was this question to answer?

   3. Provide another prompt to student. Give time to read. Let student keep this information for the rest of the interview.

Verbal Probes:

a. How difficult is this question to answer?
b. How is this question related to the previous two questions?
c. Please arrange the three questions in order of difficulty. (Give student time to arrange questions).
d. What do you understand as “difficult” when arranging these questions?

6. **Wrap up**

Thank you again for your valuable collaboration. Once more, **this interview is confidential,** you will not be identified by name and only the transcriber will listen to this tape. The transcriber is bound to confidentiality, as well.

Information sheets given to students during the interview:

Sheet 1: Entropy definition

Entropy (S) is a thermodynamic function that increases with the number of energetically equivalent ways to arrange the components of a system to achieve a particular state. It may be thought of as a measure of the dispersion of the energy in a system and it is associated with disorder or randomness at the molecular level.

Sheet 2: Second Law of Thermodynamics

We have seen that both the system and surroundings may undergo changes in entropy during a process. The sum of the entropy changes for the system and the surroundings is the entropy change for the universe:

$$\Delta S_{univ} = \Delta S_{sys} + \Delta S_{surr}$$

The Second Law of Thermodynamics says that for a process to be spontaneous as written (in the forward direction), ΔS_{univ} must be positive ($\Delta S_{univ} > 0$). Therefore, the system may undergo a decrease in entropy as long as the surroundings undergo a larger increase in entropy making the resulting ΔS_{univ} positive, and vice versa. A process for which ΔS_{univ} is negative is not spontaneous as written.

Prompts evaluated during the interview:

Self-explaining own answer, SEA

When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). However, this process is spontaneous. How do you explain this? Please be as thorough in your response as possible.

Self-explaining agreement/disagreement, EADA

When water freezes below 0 ºC, its change in entropy is negative ($\Delta Ssys < 0$). Despite this observation, your group members maintain that this process is spontaneous. Therefore, they say, no energy input from the outside is necessary to make this change happen. Do you agree with your classmates? Please explain and be as thorough in your response as possible.

Self-explaining for others, SEO

When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). However, this process is spontaneous. Explain this in writing so that a classmate of yours can use your explanation as reference when answering a similar problem. Your answer will be used by your classmate. Please be as thorough in your response as possible.

Self-explaining incorrect answer, SEIA

When water freezes below 0ºC, its change in entropy is negative ($\Delta S_{sys} < 0$). Your group members maintain that this process will not occur spontaneously. Therefore, they say, there must be an energy input from the outside to make this change happen; otherwise, water will not freeze. This stance is incorrect. What do you think led your classmates to this incorrect conclusion? Please be as thorough in your response as possible.

Appendix 3: Supplemental Data Analysis and Results

Data analysis

In addition to the textual analysis of the responses, we conducted a structural analysis. Linguistic studies have reported the structure of essays (i.e. use of types of conjunctions and length of text) vary with their summative-analytical nature (Durst, 1987). Although the responses in our case were relatively short to be considered short essays, we decided to investigate whether structural difference would be noticeable as a function of level of self-explaining.

Structural analysis of Learning Event data. For each *response*, we counted the total number of words and cohesive conjunctions (i.e. words used in text construction to connect sentences). For the total number of words, we considered symbols representing individual concepts such as *change in entropy of the system*, ΔS_{sys}, as a single unit. Other examples are: –ΔS_sys, +, →, H_2O. We tallied mathematical sentences using the same principle; therefore, the word count for an equation corresponded to the number of elements used in the mathematical sentence. The word count for the following two examples is three: $\Delta S_{sys} < 0$; ΔS_universe = 0. We tallied contractions as two words. In the case of cohesive conjunctions we used the categories shown in Table A1. Linguistic studies have shown the prevalence of causal and adversative cohesive conjunctions in analytical essays, and additive and temporal cohesive conjunctions in summative essays (Durst, 1987). We compared the mean word count by SE-Tasks using ANOVA. We calculated the ratio of each cohesive conjunction-type by dividing the frequency by the total word count. We used these cohesive conjunction-type ratios as observed variables for the subsequent Latent Profile analysis. We postulated that the overt explanatory behaviour of the students would be associated with the structural characteristics (total word count and cohesive conjunction-type ratios).

Table A1 Cohesive Conjunction Categories as Described by Durst (1987)

Cohesive Conjunction categories	Description and examples
Additive	Indicates coordination; two sentences are given equal weight. Examples include conjunctions such as "and," "also," "furthermore," "or", “plus”, “that”.
Temporal	Conjunctive relation showing chronological connection. Examples include "after," "then," "when,” “once”, “while".
Causal	Indicate cause and effect relation. Examples include "because," "so," "therefore," "thus", “since”, “due”, “as”, “if”.
Adversative	Indicate that what follows contrasts with what has just been said. Examples include "in fact," "but," "however," "instead", “although”, “whereas”, “though”, “yet”.

We performed LPA using the conjunction-type ratios and total word count as observed variables (i.e., five observed variables). The output of the LPA analysis of the conjunction-type ratios was the categorization of students into distinct profile classes based on their text construction, the Text Construction Profiles, TC-Profile. Table A2 shows examples of coded responses from the learning event data, and the corresponding observed variables we used in the latent profile analyses.

Table A2 Example Report of Textual Analysis and Structural Analysis of Learning Event Responses

<table>
<tr><th rowspan="3">Response Examples</th><th colspan="7" rowspan="2">Textual Analysis</th><th colspan="6">Structural Analysis</th></tr>
<tr><th rowspan="2">Value</th><th rowspan="2">Total words</th><th colspan="4">Cohesive Conjunctions</th></tr>
<tr><th>Value</th><th>BI</th><th>DI</th><th>E</th><th>P</th><th>U</th><th>NR</th><th>Additive</th><th>Temporal</th><th>Causal</th><th>Adversative</th></tr>
<tr><td rowspan="2">The change in entropy for the system is less than zero therefore the ΔS_surr must have a larger increase of entropy than the negative decrease of the entropy of the system for the process to be spontaneous, meaning ΔS_univ will still be greater than 0.(DI) When water freezes the system loses entropy, but the outside surroundings gain more entropy than what was lost by the system.(BI)</td><td>Count</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>Count</td><td>66</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Code--ratio</td><td>0.5</td><td>0.5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>Count per 100 words</td><td>-</td><td>0</td><td>1.52</td><td>1.52</td><td>1.52</td></tr>
<tr><td rowspan="2">Water freezes at 0˚C.(P) After frozen, no matter how much more energy is lost (E = heat) it’s still just as frozen – it can never be “more frozen” w/ more cold.(E) The process is spontaneous, yet negative because the system is more negative than the surroundings are positive.(DI)</td><td>Count</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>Count</td><td>48</td><td>0</td><td>0</td><td>2</td><td>1</td></tr>
<tr><td>Code-ratio</td><td>0</td><td>0.33</td><td>0.33</td><td>0.33</td><td>0</td><td>0</td><td>Count per 100 words</td><td>-</td><td>0</td><td>0</td><td>4.17</td><td>2.08</td></tr>
</table>

Association analysis between self-explaining tasks and self-explaining profile. We used Chi-square tests to determine the association between Text Construction Profiles membership, TC-Profile, and the Self-Explaining Task, SE-Task. We used IBM SPSS Statistics (Version 21.0.0.0) for the Chi-square tests.

Results and discussion

Pilot Study

Following we describe results corresponding to the analysis of the pilot study dataset. These results informed the implementation of the main study and also supported findings from the main study.

Code type distribution. Students' responses were coded following the coding scheme described in the methodology section to produce a tally of the codes in each response. Initially, we analysed the association between these code types and the self-explaining task, SE-Task (Table A3). For this analysis the codes "unclassifiable, U" and "non-relevant, NR" were excluded because they did not contribute valuable insight about the self-explaining behaviour. From the total count of codes it is interesting to notice that for this dataset, the "deductive inference" code presents the highest count (Table A3). This suggests that students were actively engage in the generation of inferences. It is also noticeable that the "elaboration, E" and "paraphrasing, P" codes have high number of occurrences. These E and P codes are associated with lower explanatory sophistication as they describe responses that recount information. High counts in E and P codes thus suggested that students relied heavily on recounting information when prompted to write explanations.

The Chi-square test showed no significant association between the code types and the SE-Tasks, at a 95% confidence interval, χ^2 (9, N = 269) = 15.83, p = .07. Nonetheless, inspection of Table A3 showed evidence of a trend: higher percentages (dark grey shaded cells) of the "bridging inference, BI" code are found in the self-explaining-to-others task, SEO, and self-explaining agreement/disagreement task, EADA. This finding suggests that the SEO and EADA tasks prompted students to generate more bridging inferences to link chemistry concepts (i.e., entropy and the Second Law of Thermodynamics). In the case of the code "deductive inference, DI" the higher percentage was found in the EADA task, which suggests that students generated more deductive inferences while working on this SE-task. In the case of the DI code the SEA (self-explain own answer) and SEIA (self-explain incorrect answer) tasks had moderately high percentages too, thus engaging students in the generation of inferences as well. In the case of the "elaboration, E" code, the results suggest a similar percentage in the SEO, EADE and SEA tasks, but a higher percentage in the SEIA task. Finally, the “paraphrasing, P” code presents higher percentages in the EADA and SEIA tasks, suggesting that responses on these SE-tasks were heavily composed of recounted information. In summary, these results suggest that, although not statistically significant, the code types in the students’ responses were associated with the self-explaining tasks.

Table A3 Pilot study learning event code type distribution by SE-Task

Code Type	Total	SE-Task			
		%SEO	%EADA	%SEA	%SEIA
BI	25	40	32	16	12
DI	95	17	33	26	24
E	77	25	22	21	32
P	72	14	33	18	35
U	21	14	24	38	24
NR	28	29	29	14	29

Without codes “U” and “NR”: χ^2 (9, N = 269) = 15.83, p = .07.

The results in Table A3 show the total number of code types per SE-Task, but for 80% of the students the responses were not composed of only one code. Therefore the follow-up analysis considered the code types and the total number of codes in each response. This way, we accounted for the relationship among all the code types present in each response for the categorization of student’s self-explaining behaviour.

Latent profile analysis. We used Latent Profile Analysis (LPA) to identify patterns in code-ratios (i.e., number of code type divided by total codes in response) in students’ responses. These analyses required the selection of the best model for the data. In order to make that decision the following information was used: number of profiles selected; goodness of fit indexes (Loglikelihood (LLH), Akaike’s Information Criterion (AIC), Bayesian Information Criterion (BIC), sample-size adjusted Bayesian Information Criteria (SSA–BIC)); likelihood ratio tests (Voung–Lo–Mendell–Rubin and parametric bootstrap); and homogeneity (referred to as entropy value) and cases size in each profile. Next we describe the model selection process for the pilot study.

Consistent with common practice, we explored solutions with varying numbers of profiles and selected the one that made the most sense in terms of interpretability and model fit information. We evaluated one- to ten-profile solutions models in relation to indexes of fit commonly used for this purpose (Table A4). For the information criteria Loglikelihood, LLH, the value increased as the number of profiles increased thereby indicating progressive model fitness from the model with only one profile up to the model with ten profiles. Thus, the LLH value did not provide useful information for the model selection. For the three information indexes (AIC, BIC, and SSA-BIC), lower values indicate better model fit. Our results showed that all information indexes progressively became lower as the model solution incorporated more profiles (Figure A1). This indicated that model solutions with more profiles seemed to better fit the data. In the case of the Voung–Lo–Mendell–Rubin (VLMR) test and parametric Bootstrap Likelihood Ratio Test (BLRT) the p-value reflects how significant it is to have a model with n-profiles against a model with (n-1)-profiles ("n" being the number of profiles within each model). Therefore, if the p-values are lower than .05 this means that the model solution with n-profiles is favourable over the model solution with (n-1)-profiles. The p-values for the BLRT were all lower than .05 suggesting that any model solutions are significantly better than the corresponding previous model solution, thus these results did not help in selecting a model solution. The results in Table A4 showed that only the model solutions with two-profiles and five-profiles have a p-value lower than .05 for the VLMR test, suggesting that these model solutions were favourable over the other model solutions. Finally, the last criteria for the model solution selection are the homogeneity (i.e. entropy value) of the profiles in the model and the number of cases in each profile within the model solutions. In our case both the two-profile and five-profile solution models had high homogeneity. We selected the five-profile model solution (grey shaded cells in Table A4) because larger number of profiles increased our categorization power of students. This was a judgment call based on the model fitness, parsimony and interpretability of the five-model solution.

Table A4 Goodness of fit for LPA models based on code-ratios, pilot study dataset (N=103)

Number of Profiles	Number of Parameters	LLH	AIC	BIC	SSA–BIC	p VLMR	p BLRT	Entropy	Group Sizes LT1 %	LT5 %
LP 1	8	-85	186	207	182	—	—	—	0	0
LP 2	13	-29	85	119	78	.00	.00	.99	0	0
LP 3	18	31	-26	22	-35	.49	.00	1.0	0	0
LP 4	23	89	-131	-71	-143	.20	.00	1.0	0	2
LP 5	28	138	-221	-147	-235	.03	.00	.97	0	2
LP 6	33	183	-300	-214	-318	.79	.00	.97	0	2
LP 7	38	218	-359	-259	-379	.46	.00	.95	0	2
LP 8	43	253	-420	-307	-443	.62	.00	.98	1	3
LP 9	48	276	-457	-330	-482	.85	.00	.99	0	2
LP 10	53	305	-503	-364	-531	.25	.00	.99	0	5

Note: LLH = Loglikelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SSA–BIC = sample-size adjusted Bayesian Information Criteria; p VLMR = p-value for the Voung–Lo–Mendell–Rubin likelihood ratio test for K versus K - 1 classes; p BLRT = p-value for the parametric bootstrap likelihood ratio test for K versus K - 1 classes. Group sizes refer to the number of groups with less than 1% and less than 5% of the cases, N = 103.

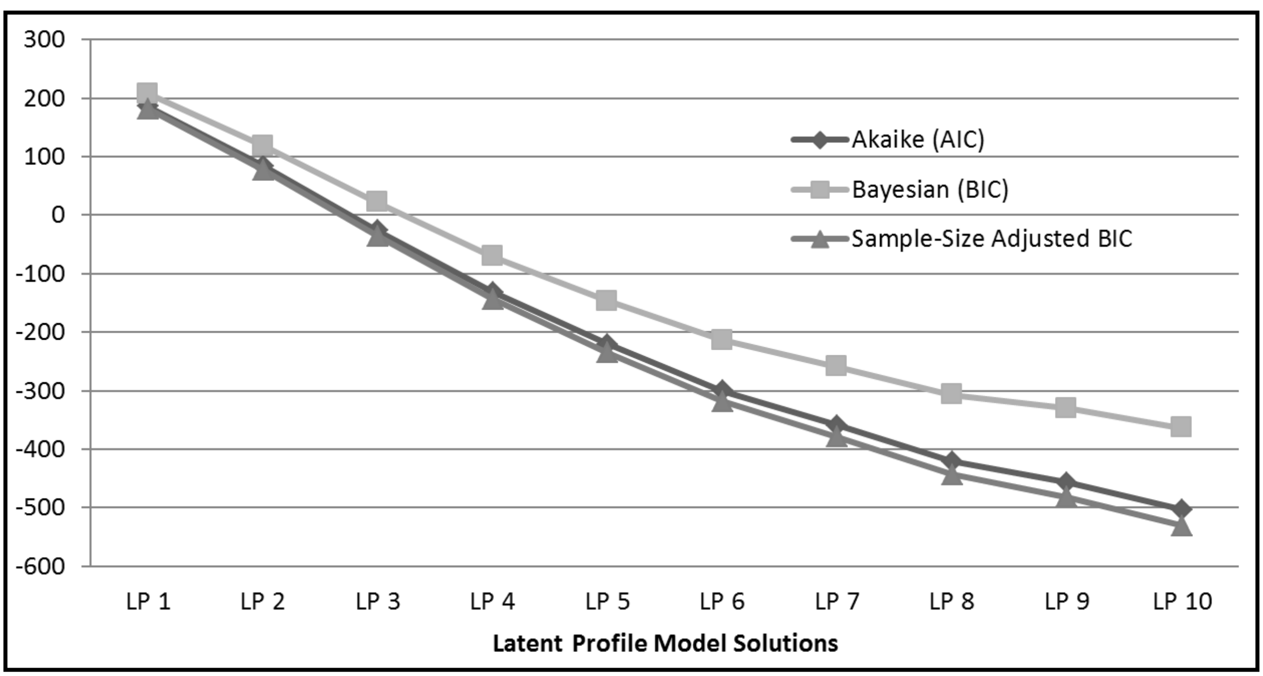

Fig. A1 AIC, BIC and Sample-size adjusted BIC values for explanation sophistication one- to ten-profile model solutions.

The profiles in the five-profile model are presented in Table A5. For each profile Table A5 shows the number of students in the group and the mean values for each of the four code-ratios. Figure A2 presents the mean values of code-ratios for each profile as a visual aid for discussion. In the case of Profile 1, this group was composed of five students whose responses only contained “bridging inference, BI” codes. Therefore, we described this profile (SE-profile) as bridging inferential. Profile 2, a three-member group, presented responses mainly using bridging and deductive inferences (89% of the response) so we described them as bridging/deductive inferential. Profile 3 is an interesting group of fourteen students whose responses used a mixture of all codes in evenly distributed ratios. We described this profile as “mixed-behaviour.” Profile 4, 63 students, had a high code-ratio for DI but also had a significant code-ratio of “paraphrasing, P”. In other words, deductive inferences predominated in these responses but students also relied significantly on recounting information. We described this profile as “deductive inferential.” Finally, the 18 students in Profile 5 relied heavily on elaboration statements, having a mean value of 92% of the response coded as “elaborations, E”. Thus, we described this profile as “elaborative.”

Table A5 Code-ratios and SE-profile descriptors for five-profile model solution (N=103)

Profile Group	**n**	**Mean code-ratio**				**SE-Profile descriptor**
		BI	**DI**	**E**	**P**	
Profile 1	5	1.00*	0.00*	0.00*	0.00*	Bridging Inferential
Profile 2	3	0.56*	0.33*	0.11	0.00	Bridging/Deductive Inferential
Profile 3	14	0.29*	0.26*	0.30*	0.15*	Mixed-behaviour
Profile 4	63	0.00	0.49*	0.15*	0.36*	Deductive Inferential
Profile 5	18	0.00	0.04	0.92*	0.04	Elaborative

* $p < .05$.

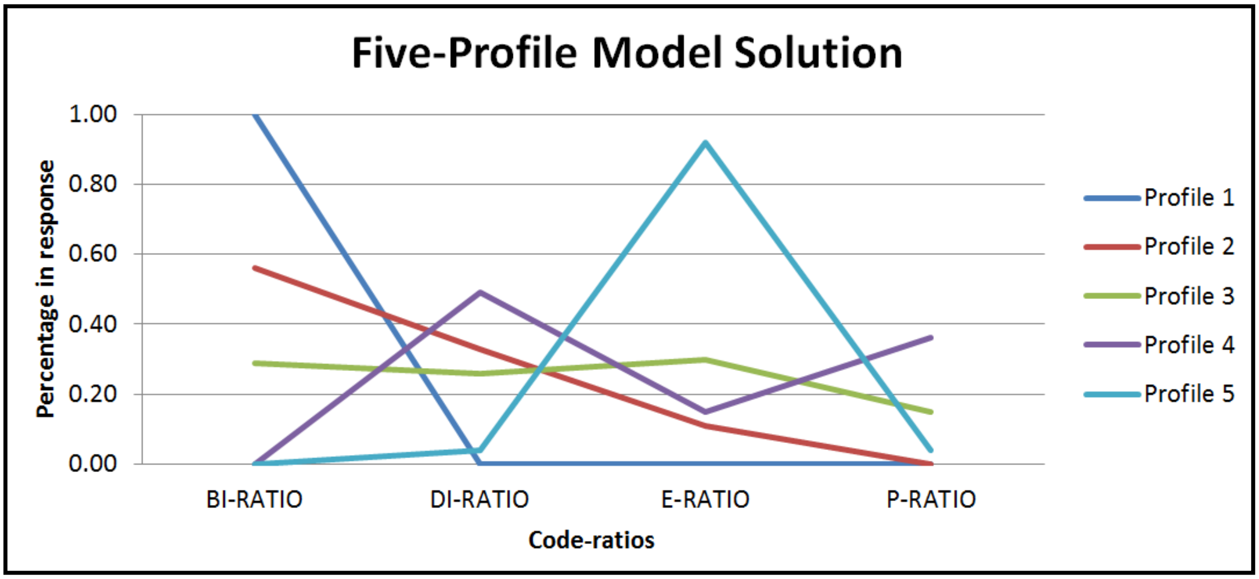

Fig. A2 Profile plot for the five-profile solution model, code-ratios.

The results in Table A5 support the interpretability of the LPA outcome (i.e., five-profile model solution) under the theoretical framework for the construct of self-explaining. This is because the LPA outcome allows the categorization of students' self-explaining behaviour in five clearly distinct groups. This finding directly addresses the first research question: Do tasks that require different self-explaining engagement induce observable categorical differences in self-explaining behaviour in the context of a General Chemistry classroom? Findings support the emergence of observable categorical differences in self-explaining behaviour when tasks prompted pupils to provide written explanations. However, to fully answer this research question we studied the association of these self-explaining behaviours (SE-Profiles) with the self-explaining task (SE-Task).

SE-Profile and SE-Task association analysis. Table A6 shows the cross tabulation of SE-Profile and SE-Task. The Chi-square test was not applicable in this case due to low sample size. This is because in the Chi-square calculation, 16 cells (80.0%) had an expected count value lower than five which is in violation of the Chi-square test requirements (less than 20% cells with expected count lower than five). Therefore the result from the Chi-square analysis was not conclusive, χ^2 (12, N = 103) = 11.69, p = .47. Nonetheless, inspection of Table A6 shows an apparent trend. SEA and SEO tasks have a higher proportion of students in the SE-Profile associated with a more analytic behaviour (i.e., bridging inferential and bridging/deductive inferential). That is, more students coming from these SE-Tasks engaged in drawing inferences and connecting ideas. Conversely, SEIA and EADA SE-tasks have higher proportions of students in the least analytic behaviours (i.e., elaborative, and deductive inferential). The apparent trend in Table A6 suggests that, although not statistically significant, the self-explaining tasks (SE-Tasks) were associated with the self-explaining behaviours (SE-Profiles).

Table A6 Percentage distribution of SE-Profile across SE-Task (N=103)

SE-Profile	n	SE-Task			
		%SEA	%SEO	%EADA	%SEIA
Bridging Inferential	5	40	40	20	-
Bridging/Deductive Inferential	3	33	33	33	-
Mixed-behaviour	14	7	36	36	21
Deductive Inferential	63	27	16	30	27
Elaborative	18	22	28	11	39

Main Study

Following we describe additional results corresponding to the analysis of the main study dataset. This information goes to a level of technical detail deeper than that discussed in the manuscript.

LPA seven-profile model solution selection. We used Latent Profile Analysis (LPA) to identify patterns in code-ratios (i.e., number of code type divided by total codes in response) in students' responses. These analyses required the selection of the best model for the data. In order to make that decision the following information was used: number of profiles selected; goodness of fit indexes (Loglikelihood (LLH), Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), sample-size adjusted Bayesian Information Criteria (SSA–BIC)); likelihood ratio tests (Voung–Lo–Mendell–Rubin and parametric bootstrap); and homogeneity (referred to as entropy value) and cases size in each profile.

Consistent with common practice, we explored solutions with varying numbers of profiles and selected the one that made the most sense in terms of interpretability and model fit information. We evaluated one- to ten-profile solutions models in relation to indexes of fit commonly used for this purpose (Table A7). For the information criteria Loglikelihood, LLH, the value increased as the number of profiles increases indicating progressive model fitness from the model with only one profile up to the model with ten profiles. Thus, the LLH value showed that a model with more profiles is favoured. For the three information indexes (AIC, BIC, and SSA-BIC), lower values indicate better model fit. Our results showed that all information indexes progressively became lower as the model solution incorporated more profiles (Figure A3). As the figure shows, the values decreased as the number of latent profiles increased up to seven and then they started to level off. This indicated that model solutions higher than seven-profiles are favoured with no much improvement after seven profiles.

In the case of the Voung–Lo–Mendell–Rubin (VLMR) test and parametric Bootstrap Likelihood Ratio Test (BLRT) the p-value reflects how significant it is to have a model with n-profiles against a model with (n-1)-profiles ("n" being the number of profiles within each model). Therefore, if the p-values are lower than .05 this means that the model solution with n-profiles is favourable over the model solution with (n-1)-profiles at a 95% confidence interval. In the case of the VLMR test the model solutions for two, three, six showed values lower than .05, also the seven-profile model solution showed a low p-value indicating that this solution is considerably good. The p-values for the BLRT were all lower than .05 suggesting that any model solutions are significantly better than the corresponding previous model solution, thus these results did not help in selecting a model solution. The entropy value for all model solution was close to the highest possible value of one, which means the homogeneity of the profiles in each solution is high, which is favourable.

Inspection of the eight- to ten-profile model solutions showed three or more group sizes with less than 5% of the total cases. We considered that the eight- to ten-profile model solutions did not add valueable insight into the categorization of the students. Based on the results from Table A7 we selected the seven-profile model solution.

Table A7 Goodness of fit for LPA models based on code-ratios, main study data (N=128)

Number of Profiles	**Number of Parameters**	**LLH**	**AIC**	**BIC**	**SSA–BIC**	***p* VLMR**	***p* BLRT**	**Entropy**	**Group Sizes**	
									LT1 %	**LT5 %**
LP 1	8	-94	205	228	202	-	-	-	0	0
LP 2	13	-41	108	145	104	.04	.00	.95	0	0
LP 3	18	21	-6.2	45	-12	.03	.00	.99	0	0
LP 4	23	86	-127	-61	-134	.26	.00	1.0	0	0
LP 5	28	153	-250	-170	-258	.36	.00	1.0	0	2
LP 6	33	199	-332	-238	-342	.01	.00	.96	0	2
LP 7	38	259	-442	-334	-454	.07	.00	.98	0	2
LP 8	43	273	-460	-337	-473	.54	.00	.98	0	3
LP 9	48	329	-562	-425	-577	.35	.00	.98	0	4
LP 10	53	343	-580	-429	-597	.60	.00	.99	0	5

Note: LLH = Loglikelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SSA–BIC = sample-size adjusted Bayesian Information Criteria; p VLMR = p value for the Voung–Lo–Mendell–Rubin likelihood ratio test for K versus K - 1 classes; p BLRT = p value for the parametric bootstrap likelihood ratio test for K versus K - 1 classes. Group sizes refer to the number of groups with less than 1% and less than 5% of the cases, N = 128.

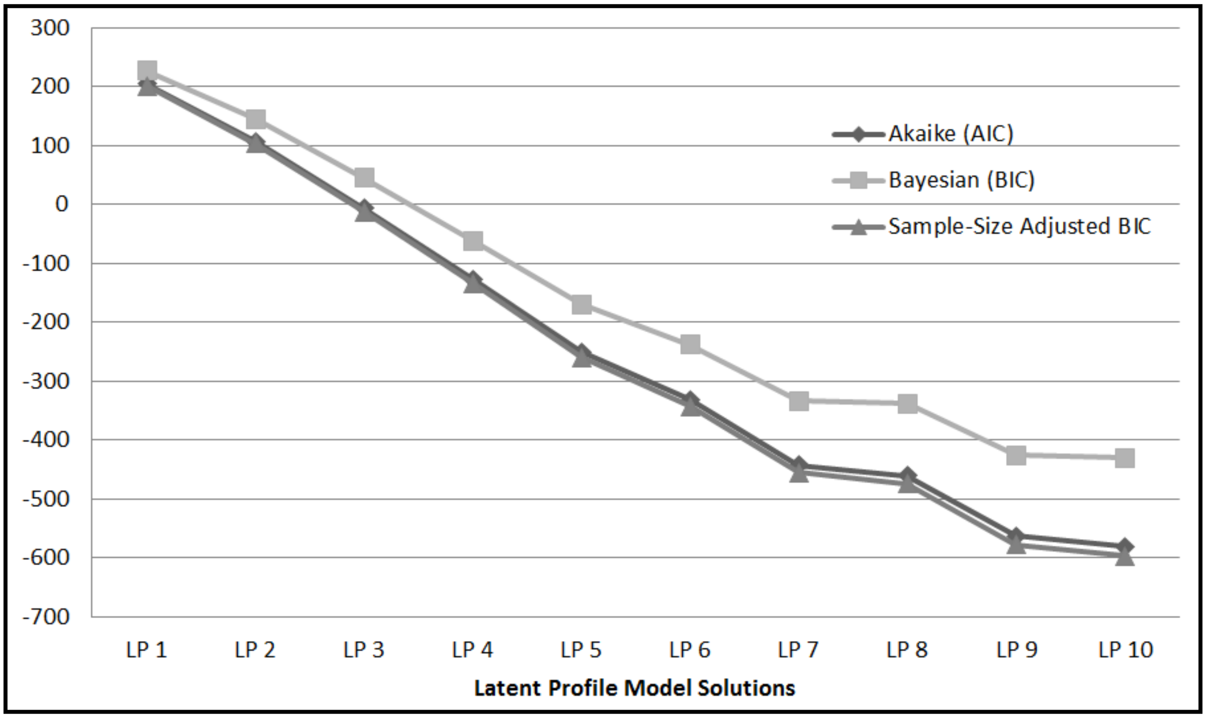

Fig. A3 AIC, BIC and Sample-size adjusted BIC values for explanation sophistication one- to ten-profile solutions.

Structural Analysis. We analysed the responses of the students in the main study phase for their structural composition in terms of (1) the total word count and (2) the cohesive conjunction type count (Table A8). Analysis showed no relevant differences among the self-explaining tasks, SE-Task, and self-explaining profiles, SE-Profiles, across these two counts. These results suggest that the text construction of the written responses is not different between participants doing different self-explaining tasks, or behaving differently when self-explaining. However, we acknowledge that the extension of the participants' responses is not as extensive as in the case of other research studies (e.g., mean total words = 500) (Durst, 1987), and this may impact our resolution.

Table A8 Descriptive statistics of word counts by SE-Task (N=128)[a]

SE-Task	n	Total words Mean (SD)	Cohesive conjunctions per 100 words Mean (SD)			
			Additive	**Temporal**	**Causal**	**Adversative**
SEA	29	61 (24)	2.4 (2.0)	1.5 (1.7)	4.3 (2.7)	0.72 (0.97)
EADA	31	63 (23)	2.3 (2.0)	0.9 (1.1)	3.7 (2.5)	0.42 (0.85)
SEO	35	55 (24)	2.1 (2.0)	1.3 (1.6)	4.0 (3.1)	0.49 (0.77)
SEIA	33	64 (19)	3.0 (2.6)	1.1 (1.1)	2.6 (2.1)	0.76 (1.12)
Total	128	61 (22)	2.2 (2.2)	1.2 (1.4)	3.6 (2.7)	0.60 (0.94)
ANOVA	*F*	1.17	2.03	.96	2.68	1.00
(3, 124)	*p*	.33	.11	.41	.05	.40
SE-Profile	**n**	**Total words Mean (SD)**	**Cohesive conjunctions per 100 words Mean (SD)**			
			Additive	**Temporal**	**Causal**	**Adversative**
Bridging Inferential	25	60 (21)	2.5 (2.3)	1.3 (1.6)	4.0 (3.3)	0.60 (0.78)
Mixed behavior	12	66 (22)	3.2 (2.8)	1.2 (1.4)	2.4 (1.8)	0.66 (1.14)
Deductive Inferential	20	46 (24)	4.2 (3.8)	1.5 (1.8)	4.3 (3.1)	0.42 (0.96)
Elaborative	24	62 (19)	3.5 (2.4)	1.4 (1.2)	3.3 (1.9)	0.67 (0.98)
Summative	47	63 (22)	3.4 (2.4)	1.0 (1.2)	3.7 (2.7)	0.64 (0.99)
Total	128	59.9 (22)	3.4 (2.7)	1.22 (1.4)	3.7 (2.7)	0.60 (0.95)
ANOVA	*F*	2.65	1.3	0.65	1.24	.24
(3, 124)	*p*	.04	.28	.63	.30	.92

[a] Six responses were unintelligible and therefore removed from the analysis.

To further analyse these data we used LPA to investigate categorical differences among students' text construction behaviours, TC-Profiles. The idea of this LPA study was to identify groups of students with similar text construction styles (in terms of the length of the explanation and the used of cohesive conjunction words) when writing explanations. In contrast with the previously shown analysis of variance, ANOVA, which used only one word count in each analysis, the LPA used all of the five different word count values in each response as observed variables to classified students into text construction profiles, TC-Profiles. This allowed an in-depth categorization of students' text construction behaviours. Next we studied potential differences of the text construction styles (TC-Profiles) across the experimental conditions (SE-Tasks) and self-explaining behaviours (SE-Profiles).

Latent profile analysis: Text construction profiles. The selection process for the best latent profile fit model solution followed the procedures described previously (see above). Results for the goodness of fit indexes are shown in Table A9 and Figure A4. The three-profile model solution showed the best fit for the data. Also the high value of homogenity (i.e., Entropy value = .998) in the three-profile model solution suggested that students' membership within each of the three profiles was well established. This meant that all students within each profile had low uncertainty of belonging to other profile within the model solution.

Table A9 Goodness of fit for LPA models based on word counts, main study data (N=128)

Number of Profiles	Number of Parameters	LLH	AIC	BIC	SSA–BIC	*p* VLMR	*p* BLRT	Entropy	Group Sizes	
									LT1 %	LT5 %
LP 1	10	-1566	3151	3180	3148	-	-	-	0	0
LP 2	16	-1537	3105	3151	3100	.55	.00	.94	0	0
LP 3	22	-1479	3002	3065	2996	.04	.00	1.0	0	1
LP 4	28	-1434	2923	3003	2915	.50	.00	.99	0	1
LP 5	34	-1412	2893	2990	2882	.60	.00	.98	0	1
LP 6	40	-1386	2851	2965	2839	.26	.00	.99	1	1
LP 7	46	-1365	2823	2954	2808	.13	.00	.96	1	1
LP 8	52	-1348	2801	2949	2785	.35	.00	.94	1	1
LP 9	58	-1331	2779	2944	2761	.81	.60	.94	1	2
LP 10	64	-1305	2738	2921	2718	.81	.02	.95	1	3

Note: LLH = Loglikelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SSA–BIC = sample-size adjusted Bayesian Information Criteria; p VLMR = p value for the Voung–Lo–Mendell–Rubin likelihood ratio test for K versus K - 1 classes; p BLRT = p value for the parametric bootstrap likelihood ratio test for K versus K - 1 classes. Group sizes refer to the number of groups with less than 1% and less than 5% of the cases, N = 128.

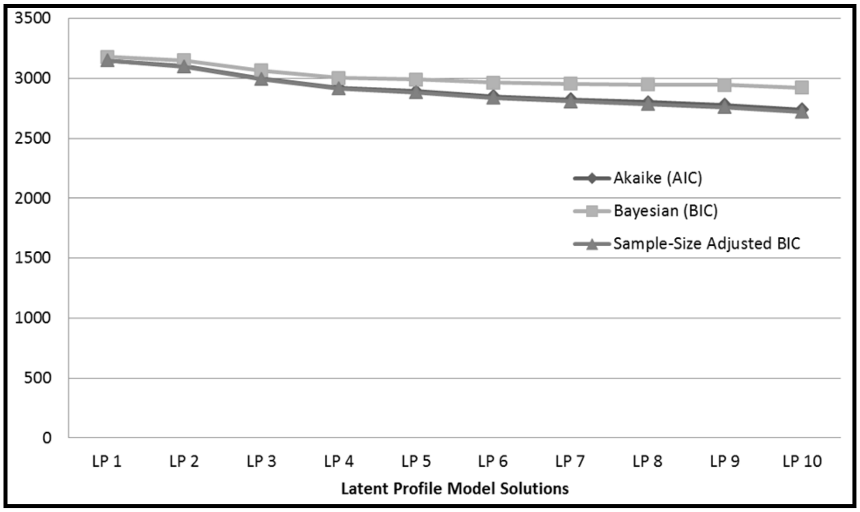

Fig. A4 AIC, BIC and Sample-size adjusted BIC values for text construction one- to ten-profile model solutions.

The results for the three-profile solution are shown in Table A10. Most profiles presented substantial differentiation among word counts (i.e., total word and cohesive conjunction counts). Figure A5 shows the cohesive conjunction counts for each profile as visual aid. Profile 1 showed the highest count of adversative cohesive conjunctions of the three profiles. We described this profile as "Adversative." This profile size is small in comparison with the other and for practical reasons subsequent analyses did not consider it. Profile 2 showed the highest count of causal cohesive conjunctions, therefore we described it as "Causal." Finally, Profile 3 showed the highest number of total word count. We described this profile as "longer-texts."

Table A10 Word counts for text construction three-profile model solution

Profile	**n**	**Total words Mean (SD)**	**Cohesive conjunction per 100 words Mean (SD)**				**TC-Profile Descriptor**
			Additive	**Temporal**	**Causal**	**Adversative**	
Profile 1	4	63 (21)*	1.3 (2.2)*	0.8 (1.4)*	2.2 (2.6)*	3.7 (0.3)*	Adversative
Profile 2	84	54 (21)*	2.2 (2.2)*	1.3 (1.4)*	4.0 (2.6)*	0.0 (0.3)	Causal
Profile 3	40	74 (21)*	2.5 (2.2)*	1.1 (1.4)*	2.9 (2.6)*	1.5 (0.3)*	Longer-texts

* $p < .05$.

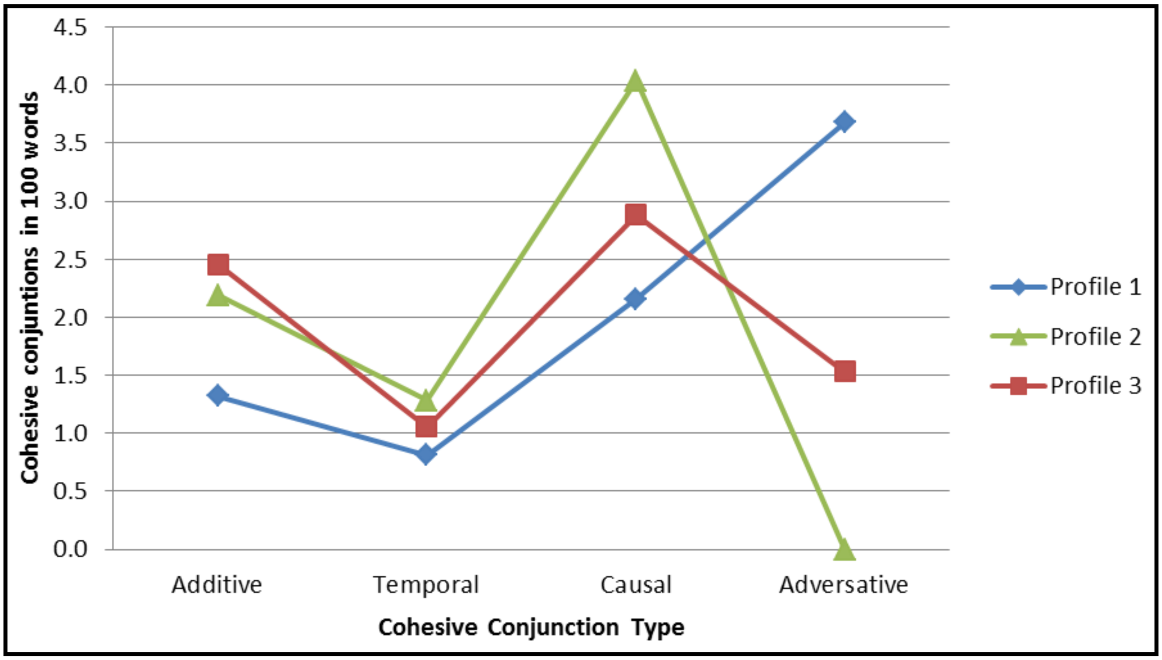

Fig. A5 Profile plot for the three-profile solution model, cohesive conjunctions in 100 words.

TC-Profile and SE-Task association analysis. For the association analysis of the TC-Profiles and the SE-Task, we did not consider the "adversative" profile due to its low number of cases. The Chi-square test showed no significant association between the remaining two TC-Profiles and SE-Tasks at a 95% confidence interval, $\chi^2(3, N = 124) = 3.30, p = .35$ (Table A11). This result suggests that the text construction behaviour of the students was not found to be different depending of the self-explaining task. This finding provides further support to the previously discussed ANOVA study (see above) as no significant differences were found among the writing styles of the students across the SE-Tasks.

Table A11 Percentage distribution of TC-Profile across SE-Task[a]

TC-Profile	n	SE-Task			
		%SEA	%EADA	%SEO	%SEIA
Adversative	4	25	25	-	50
Causal	40	28	15	28	30
Longer-texts	84	20	29	29	23

[a] Without "Adversative" profile: χ^2 (3, $N = 124$) = 3.30, $p = .35$.

SE-Profile and TC-Profile association analysis. As in the previous analysis, for the association analysis of the SE-Profiles and the TC-Profiles, we did not consider the “adversative” profile due to its low number of cases. The Chi-square test showed no significant association between the remaining two TC-Profiles and SE-Profile at a 95% confidence interval, $\chi^2(4, N = 124) = 3.37$, $p = .50$ (Table A2). This result suggests that the self-explaining behaviour of the student is not significantly associated to students’ text construction behaviour in terms of use of cohesive conjunction types and text extension. We acknowledge that the low mean values for total words in the students’ response presents a limitation for the resolution and power of this result.

Table A12 Percentage distribution of SE-Profile across TC-Profile[a]

SE-Profile		**TC-Profile**		
	n	**%Adversative**	**%Causal**	**%Longer-text**
Bridging Inferential	25	-	60	40
Mixed behaviour	12	8	67	25
Deductive Inferential	20	5	80	15
Elaborative	24	4	63	33
Summative	47	2	64	34

[a] Without “Adversative” profile: χ^2 (4, $N = 124$) = 3.37, $p = .50$.