Annotation of empathy in student-written peer reviews

May 2020

Table of Contents

List of Figures

1. Introduction

Nowadays, most information is readily available to people and solely reproduction of information is losing attention. This manifests in a shift of job profiles towards interdisciplinary, ambiguous and creative tasks (vom Brocke et al. 2018). Therefore, educational institutions need to evolve in their curricula, especially regarding the compositions of skills and knowledge conveyed. Particularly teaching higher order thinking skills to students, such as critical thinking, collaboration or problem-solving, has become more important during the last few years (Fadel et al. 2015). This has already been recognized by the Organization for Economic Co-operation and Development (OECD), which included these skills as a major element of their Learning Framework 2030 (OECD 2018). One elementary skill for communication and successful team work represents the "*ability to simply understand the other person's perspective […] and to act emotionally on the other",* also defined as empathy (Davis 1983). However, studies have shown that empathy skills of students have decreased from 1979 to 2009 by more than thirty percent and even more rapid in the last period of the study from 2000 to 2009 (Konrath et al. 2011).

Therefore, our aim is to create an adaptive empathy learning tool, that supports student with a learning environment to improve being more empathetic. By creating a writing support interface, students will get instant feedback from a pre-trained algorithm on their degree of empathy in written texts, e.g., when writing a peer-review on a fellow student's business idea. However, in order to leverage recent methods from Natural Language Processing and Machine Learning, a high-quality annotated corpus is needed to train a predictive model. This guideline helps to evaluate the corpus on empathy and foster a shared understanding on how empathy in student-written peer reviews can be detected.

2. Empathy in a nutshell

The ability to perceive the feelings of another person and to react to their emotions in the right way requires empathy – the ability *"of one individual to react to the observed experiences of another"* (Davis 1983, p.1). Empathy plays an essential role in daily life in many practical situations, such as client communication, leadership or agile teamwork. Therefore, especially business schools today are increasingly trying to focus on fostering empathy skills (e.g., Peterson and Limbu 2009) to provide students with the right skill set to meet future job profiles (i.e., vom Brocke et al. 2018). The importance of empathy and other metacognition skills has been manifested by the *OECD*, which included them as a major element of their *Learning Framework 2030* (OECD 2018). Despite the interdisciplinary research interest, the term *empathy* is defined from multiple perspectives in terms of its dimensions or components (Decety and Jackson 2004). Empathy can be divided into various categories and subscales. Davis' (1983) proposed in his studies the four scales *fantasy scale* (imaginatively transpose oneself into fictional situations), *perspective taking* (ability to shift perspectives), *empathic concern* (degree to which the respondent experiences feelings of warmth, compassion and concern for the observed individual) and *personal distress* (individual's own feelings of fear, apprehension and discomfort at witnessing the negative experiences of others). Other authors and today's widely accepted distinction of empathy distinguishes between emotional (affective) and cognitive empathy, whereas emotional empathy lets us feel what others are feeling and cognitive empathy is the human's ability to recognize and understand others. Being aware that there are multiple perspectives on empathy, in this paper we focus on the *cognitive and emotional components* of empathy as defined by Davis (1983) and Lawrence et al. (2004). Therefore, we follow the *"Toronto Empathy Scale" (*Spreng et al. 2009) as a synthesis of instruments for measuring and validating empathy. Empathy refers to the "*ability to simply understand the other person's perspective […] and to react to the observed experiences of another" (Davis 1983, p.1),* where empathy consists of both *emotional and cognitive components* (Spreng et al. 2009). While emotional empathy lets

us perceive what other people feel, cognitive empathy is the human ability to recognize and understand other individuals (Lawrence et al. 2004).

3. Data domain: student-written peer review

This annotation study is conducted on a set of German student-written peer reviews. The data was collected throughout a mandatory course of the master's program in Business Innovation at the University of St. Gallen. In this course, the students were asked to develop and present a new business model. Each student then received three different peer reviews, in which a fellow student from the same course elaborated on the strengths and weaknesses of the business model and gave persuasive recommendations and suggestions for improvement. The reviews were submitted online through a learning platform. The dataset for this annotation study contains a random subset of 500 peer reviews, collected from more than 7,000 documents over the last few years.

4. Empathy in student-written peer reviews

This chapter gives concrete guidelines on how to define empathy in the given peer review dataset and aims to establish a shared understanding of empathy in review texts. According to the elaboration of empathy mentioned in the first chapter, both approaches and scales are taken into consideration. However, since the reviews are evaluated on activities based on a new business idea of the student, Davis' fantasy scale and personal distress do not match. The fantasy scale denotes the tendency to transpose oneself into fictional characters in books, movies or plays. Since the data domain is about real-life business models, it does not represent such an environment of fictional characters. Additionally, students did not express personal negative experiences in their business models but rather present their business model in a logical, factual, and convincing manner. Thus, the scale of personal distress will not be included either. This leaves us with Davis' perspective taking and empathetic concern, as well as with cognitive and emotional empathy. Both approaches can be put together and applied to the context of peer reviews according the following:

- *Cognitive empathy (perspective taking):* The students use cognitive processes such as role taking, perspective taking or "decentering"[1] while evaluating the peers' submitted tasks. This means students set aside their own perspective and "step into the shoes of the other". Cognitive empathy can happen purely cognitive in that there is no reference to any affective state (Baron-Cohen and Wheelwright 2004), but mostly includes understanding the other's emotional state as well. The following extract from a student-written peer review, translated to English, demonstrates high cognitive empathy: *"You could then say, for example, 'Since market services are not differentiated according to customer segments and locations, the following business areas result... And that due to the given scope of this ITPA you will focus on the Concierge-Service business segment.' After that, you have correctly only dealt with this business segment."* When annotating, it is helpful to read the statements from the feedback-receiver's perspective and think about if the reviewer is trying to help you with your further elaboration of your business idea and if he/she truly tries to put himself/herself in your shoes and find important hints, thoughts or comments.

- *Emotional empathy (empathic concern):* The students respond emotionally to the peers' affective state. The students can either show the same emotions as read in the review or simply state an appropriate feeling towards the peer. Typical examples include sharing excitement with the peer about the business model submitted or showing concern towards the peer's opinion The following extract from a student-written peer review, translated to English, shows high emotional empathy: *"I think your idea is brilliant!"* When annotating, it is helpful to read the statement from the feedback-writer's perspective and think about if he/she managed to respond emotionally to the peer's business idea by showing excitement, concern, disbelief, etc.

[1] Responding nonegocentrally (Piaget, 1932).

Each element of empathy will be evaluated on a scale from 1-5 (see figure 1).
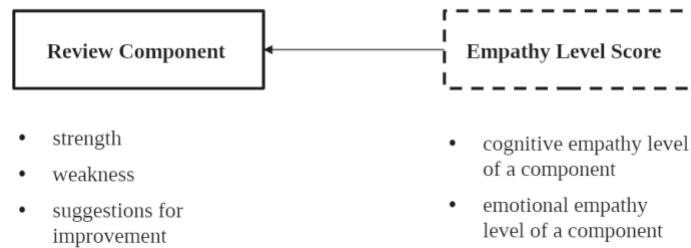

Figure 1: Annotation scheme (own illustration)

The differences between cognitive and emotional empathy should always be kept in mind. Statements can both be similar or very different in regards of emotional and cognitive empathy. The following example shows a component with high cognitive, but low emotional score: *Vielleicht noch ein wenig mehr ins Detail gehen und sich überlegen, wie die einzelnen Funktionen aussehen. Am Anfang ein paar weniger Funktionen anbieten und genauer auf einen Kundenwunsch eingehen. Ist das Kernprodukt, dass ich etwas mit meiner Freundin zusammen planen kann, oder der Aspekt es danach zu Veröffentlichen oder ist es einfach nur ein Tool, mit dem ich einfacher meine Daten zu meinem Urlaub sortieren kann? Ausgangs Pain ist ja die Überforderung mit zu viel Information. Wie genau behebst du das? Du fügst ja eher noch eine Informationsebene dazu indem man sich zusätzlich noch Boards von Freunden anschauen kann..*

The student did manage to put himself/herself into the peer's perspective. However, he/she did hardly show personal emotions.

The same concept can be applied for the other case. The following example illustrates a high emotional, but low cognitive score: *Ich finde deine Idee wirklich sehr, sehr gut!*

The student did manage to show a lot of excitement towards the peer's business idea. However, since it is missing any further explanations or supporting sentences (e. g. what particular is good about the idea), it receives a low cognitive score.

The following tables includes more details about the cognitive and emotional empathy scores. Because the general assessment of cognitive empathy of strengths, weaknesses and suggestions for improvement varies between the components, each component is defined specifically. This is not necessary for the assessment of emotional (affective) empathy due to the possibility to generalize the evaluation criteria.

	Cognitive empathy
1 = absolutely weak	The student's review is very short and does not include the peer's perspective. *Strengths*: The student only mentions one strength. This might not be relevant at all and lacks any further explanation, detail or example. *Weakness:* The student only mentions one weakness. This might not be relevant at all and lacks any further explanation, detail or example. *Suggestions for improvement:* The student only mentions one suggestion. The suggestion is not followed by any explanation or example and might not be relevant for the further revision of the peer.
2 = very weak	The student did not try to understand the peer's perspective. The student rather just tried to accomplish the task of giving feedback. *Strengths:* The student mentions one or more strengths. They could be relevant for the peer. However, he does not add any further explanation or details. *Weaknesses:* The student states one or more weaknesses without explaining why they are seen as such. They could be relevant for the peer. However, the statements do not include any further elaboration on the mentioned weakness. *Suggestions for improvement:* The student suggests one or more improvements that could be relevant for the peer. However, the student does not explain why he/she suggests the change or how the suggestions for improvement could be implemented.
3 = slightly weak / equal	The student tries to understand the perspective of the peer and adds further elaborations on his statements. However, his elaborations are not completely thought-through and his feedback is missing some essential explanations, examples, or questions to make sure he/she understood right. *Strengths:* The student mentions one or more strengths and explains some of them with minor explanations or examples on why it is seen as a strength. However, most strengths focus on formal aspects rather than contextual aspects. *Weaknesses:* The student states one or more weaknesses and explains some of them with minor explanations or examples. The student could also just state questions to illustrate the weakness in the peer's business idea. Most weaknesses are not explained why they are such. *Suggestions from improvements:* The student suggests one or more improvements that are mostly relevant for the further establishment of the activity. The suggestions are written only on a high-level and most of them do not include further explanations or examples. The student explains only occasionally *why* he/she suggests a change or how it could be implemented.

4 = Fairly strong	The student thinks from the perspective of the peer. He/She elaborates in a way that serves best the peer to further establish the idea or activity. Each component is affirmed with further explanations. *Strengths:* The student was able to recognize one or more strengths that are helpful for the peer to affirm their business idea and activity. He/She highlights contextual strengths rather than formal strengths. The student supports most statements with examples or further personal thoughts on the topic but might still be missing some reasonings. *Weaknesses:* The student thinks from the peer's perspective and what would help him/her to further succeed with the task. This could be demonstrated by stating various questions and establishing further thoughts. The student explains the weakness and adds examples, but he/she is still missing some reasonings. *Suggestions for improvement:* The student suggests one or more improvements that are relevant for the further establishment of the activity and idea from the perspective of the peer. Most suggestions are written concrete and, if applicable, supported by examples. In most cases, the student explains *why* he/she suggests a change.
5 = strong	The student fully understands the peer's thoughts. He/She completely stepped outside his/her own perspective and thinks from the peer's perspective. He/she does that by carefully evaluating the peer's idea according to its strengths, weaknesses and suggestions for improvement. Questions, personal pronouns or direct addressing of the author could be used in order to better understand and elaborate on the peer's perspective. *Strengths:* The student fully grasps the idea of the peer. He/She elaborates on strengths that are important for the peer for his continuation of the task and adds explanations, thoughts or examples to his statements, reasoning why the strength is important for the business idea. *Weaknesses:* The student thinks completely from the peer's perspective and what would help him/her to further succeed with the task. The student explains the weakness in a very detailed manner and describes why the weakness is important to consider. He could also give counterarguments or ask questions to illustrate the weakness. *Suggestions for improvement:* The student suggests improvements as if he would be in the peer's perspective in creating the best possible solution. The student completes his suggestions with rich explanations on *why* he/she would do so and elaborates on the improvements in a very concrete and detailed way. Almost every suggestion is supported by further explanations.

Examples (Suggestions for improvement)

1. Auf der 2ten Slide in der Legende zwischen Gast und Restaurant. fehlt ein d, das muss hinzugefügt werden.
   - *Only one statement*
   - *Not relevant for the further elaboration of the peer's business idea and tasks*
   - *Written objectively and without personal pronouns*
2. Ich würde noch überlegen, ob du die Kunden besser einbeziehen kannst. Zudem musst du das BPMN nochmals überarbeiten.
   - *Two statements, they could be relevant for the peer*
   - *No further explanations or examples, why the peer should consider doing these changes*
3. Konkret würde ich am Anfang zwei Aufgaben nehmen Registrierung, Kreditanfrage und von Letzterer aus eine Nachricht an die Plattform senden. Da erscheint ein NachrichtenSymbol das in zwei aufeinanderfolgenden automatischen Aufgaben erst Vorprüfung, dann Vorauswahl mündet.

Anschliessend ein exklusives Gateway mit zwei Pfeilen positiv, negativ. Negativ Aufgabe Absage versenden über eine EndeNachricht dunkler Kreis wieder zum Kreditnehmer in die Aufgabe Bescheid erhalten. Positiv Pfeil zu automatischer Aufgabe mit Aufforderung Daten zur Verfügung stellen und Nachricht in die Aufgabe beim Kreditnehmer Bescheid erhalten. Beim Kreditnehmer Swimlane würde ich ein Gateway mit zwei Pfeilen Anfrage abgelehnt, Anfrage genehmigt. Bei Ablehnung erfolgt das EndeSymbol. Bei einer Genehmigung folgt die Aufgabe Zusätzliche Daten auf Plattform laden. Danach erfolgt beim Kreditnehmer nur noch die Aufgabe Bescheid erhalten. Bei der Plattform folgt die Aufgabe Konsolidierung der Daten Durch Sequenzfluss folgend auf die Nachricht nach der Aufgabe Aufforderung Daten zur Verfügung stellen. Danach die beiden manuellen Aufgaben aus der Aufgabenstellung und die Aufgabe Freischaltung auf der Plattform. Nun stosst die Swimlane des Anlegers dazu. Ich würde eine Schleifenaufgabe bei Registrierung der Zusagen einbauen.

- *More than two statements that are mostly relevant*
- *Some suggestions are supported by further explanation or example*
- *However, most statements are missing further elaboration on why the suggestions is made*

4. Vielleicht noch ein wenig mehr ins Detail gehen und sich überlegen, wie die einzelnen Funktionen aussehen. Am Anfang ein paar weniger Funktionen anbieten und genauer auf einen Kundenwunsch eingehen. Ist das Kernprodukt, dass ich etwas mit meiner Freundin zusammen planen kann, oder der Aspekt es danach zu Veröffentlichen oder ist es einfach nur ein Tool, mit dem ich einfacher meine Daten zu meinem Urlaub sortieren kann? Ausgangs Pain ist ja die Überforderung mit zu viel Information. Wie genau behebst du das? Du fügst ja eher noch eine Informationsebene dazu indem man sich zusätzlich noch Boards von Freunden anschauen kann...

- *Several statements that are relevant for the further elaboration on the peer's tasks*
- *Most statements are supported by further elaborations (e. g. specific questions to trigger more thoughts on the topic)*
- *Most statements are explained why the are suggested (e. g. by showing a fact that has been missed)*
- *Some elaborations could be written more concrete*

5. Ein nächster Schritt wäre eine Analyse der tatsächlichen Kaufkraft der Kunden und Anzahl potentieller Benutzer zu machen. Vor allem ist hierbei wichtig herauszufinden wie viele Bestellungen/Benutzer benötigt sind um Profit zu machen. Um das Netzwerk weiter auszubauen wird empfohlen mit Partnerorganisationen, wie Sportvereine, Vegane/Vegetarische Hersteller, und weitere zusammenzuarbeiten. Dadurch entsteht eine breite Produktpalette, die kundenspezifisch zugeschnitten werden kann. Damit eine erfolgreiche Zusammenarbeit zwischen Hersteller und Kunden generiert wird, wird empfohlen nochmals tiefgründig über ihre Beziehung zu gehen und weitere Möglichkeiten der Zusammenarbeit aufzuzeigen. Die Kundendaten allein generieren bereits Wert für den Hersteller, jedoch würde eine Art Innovation Lab, mit Meetings zwischen Herstellern und Kunden die Beziehung beispielsweise stärker gestalten. Hersteller können hierbei mit Hilfe von Kunden neue Produkte entwickeln, und diese dann von Kunden direkt testen lassen. Um die Abonnement Gebühr für Kunden so tief wie möglich zu gestalten, könnten Rabattaktionen erarbeitet oder mit Gutscheinen geworben werden. Zudem könnte man für jede Weiterempfehlung an neue Kunden eine Box gratis zur Verfügung stellen. In der Lösung besteht viel Potential was weiter erarbeitet werden müsse. Nebst der ZielgruppenAnalyse wird empfohlen, sich tiefgründig mit dem Produktangebot, sowie Pricing beschäftigen. Zudem wird empfohlen, die Beziehung zu Herstellern und zwischen Herstellern genauer zu definieren und Lösungen aufzuzeigen, um Konkurrenz innerhalb der Gruppen zu vermeiden.

- *Several statements that are all relevant for the peer's idea*

- *Statements are supported by rich explanations and further details*
- *Suggestions are explained why they are suggested and why they should be considered*

Emotional (affective) empathy	
1 = absolutely weak	The student does not respond emotionally to the peer's work at all. He/She does not show his/her feelings towards the peer and writes objectively (e.g. no "I feel", "personally" "I find this.." and no emotions such as "good", "great", "fantastic", "concerned", etc.). Typical examples would be "add a picture." or "the value gap XY is missing."
2 = very weak	Mostly, the student does not respond emotionally to the peer's work. Only very minor and weak emotions or personal emotional statements are integrated. The student writes mostly objectively (e. g. "okay", "this should be added", "the task was done correctly", etc.). In comparison to 1, he/she might be using modal verbs (might, could, etc.) or words to show insecurity in her review (rather, maybe, possibly)
3 = slightly weak / equal	The student *occasionally* includes emotions or personal emotional statements to the peer review. They could be quite strong. However, the student's review is missing personal pronouns ("I", "You") and is mostly written in third person. Emotions can both be positive or negative. Negative emotions can be demonstrated with concern, missing understanding or insecurity (e. g. with modal verbs or words such as rather, perhaps). Typically, scale 3 includes phrases such as "it's important", "the idea is very good", "the idea is comprehensible", "it would make sense", "the task was done very nicely", "It could probably be, that", etc.
4 = Fairly strong	The student was able to respond emotionally to the peer's submitted activity with suitable emotions (positive or negative). He/She returns emotions in his/her review on *various* locations and expresses his/her feelings by using the personal pronoun ("I", "You"). Some sentences might include exclamations marks (!). Typical reviews in this category include phrases such as "I am excited", "this is very good!", "I am impressed by your idea", "I feel concerned about", "I find this very..", "In my opinion", "Unfortunately, I do not understand", "I am very challenged by your submission", "I am missing", "You did a very good job", etc.
5 = strong	The student was able to respond very emotionally to the peer's work and fully represents the affectional state in his/her *entire* review. He/She illustrates this by writing in a very emotional and personal manner and expressing his/her feelings (positive or negative) throughout the review. Strong expressions include exclamation marks (!). Typical reviews in this category include phrases such as "brilliant!", "fantastic", "excellent", "I am totally on the same page as you", "I am very convinced", "personally, I find this very important, too", "I am very unsure", "I find this critical", " I am very sure you feel", "This is compelling for me" etc.

Examples (Suggestions for improvement)

1. Die USP des Konzeptes besser herausarbeiten und zeigen inwiefern sich dieses Konzept von den bisherigen Vermittlungsbüros unterscheidet.
    - *The student uses no personal emotions*
    - *Very objectively and "dry"*
2. Der Autor sollte sich nochmals genau mit den einzelnen Punkten des BMN auseinandersetzen und sich überlegen, wie die Geschäftsidee von einem Unternehmen umgesetzt/implementiert werden könnte.
    - *Rather factually*
    - *The student uses modal verbs (e. g. "der Autor sollte.")*
3. Ich würde noch auf die Schreibweise achten, damit dein tolles Beispiel nicht untergeht. Beispielsweise wurde beim Punkt Marktleistungen das Wort eine doppelt genannt. Der Abschnitt Kurz Charakteristika

Ihres Unternehmens sollte vielleicht noch einmal überarbeitet werden, da noch sehr viele grammatikalische Fehler bestehen und somit der Lesefluss gehindert wird.

- *The student occasionally illustrates emotions by using modal verbs and certain emotional expressions ("toll", "vielleicht")*
- *The student only occasionally includes personal pronouns ("Ich würde"), but writes mostly in third person ("Beispielsweise wurde", ""sollte vielleicht nochmal überarbeitet werden".*

4. Bieten alle Skigebiete genügend Empfang? Als regelmässiger Skifahrer musste ich schon einige Male erfahren, dass es viele Funklöcher gibt. Konkretisiere, falls möglich, wie man mit diesem Problem umgehen kann. Wäre es nicht spannend, mit den Skiausrüstern an den Talstationen z.B. Intersport eine Partnerschaft anzustreben? Intersport bietet die Hardware, On the Top bietet die Software. Für Intersport ein super Deal, da die ein solches Angebot wahrscheinlich auch in fünf Jahren noch nicht hinkriegen würden. Bei den Konkurrenten wären möglicherweise noch andere App-Anbieter zu beachten, die ähnliche Angebote auf den Markt gebracht haben. Du erwähnst das teure Bergsport-Angebot in der Schweiz, womit du absolut recht hast. Ich nehme jedoch an, dass auch On the Top sich noch ein Stück vom Kuchen sichern will. Frage Wie kann On the Top Geld verdienen und es dem Kunden erlauben, günstiger Skifahren zu gehen? Oder strebt das Unternehmen eher einen Added Value an?

- *The student shows personal emotions on various locations ("musste ich schon einige Male erfahren" shows annoyance, "wäre spannend» shows excitement, «womit du absolut recht hast» shows agreement)*
- *The student writes subjectively ("musste ich erfahren", "ich nehme jedoch an")*

5. Du musst zwingend eine Tabelle erstellen, welche die verschiedenen Geschäftsfeldkombinationen darstellt und dich dann auf eine beschränken. Die nachfolgenden Kapitel richten sich danach explizit nach diesem Geschäftsfeld aus. Auf S.2122 im Skript von Österle siehst du, wie du genau vorgehen sollst. Das Kapitel 7 Qualitative Beschreibung sollte ebenfalls dem Beispiel von Österle folgen. Du hast hierbei einige Unterpunkte vergessen zu erwähnen. Auch wenn deiner Ansicht nach bspw. keine Lieferanten vorhanden sind, solltest du das meiner Meinung nach zum Verständnis doch auch erwähnen. Ich verstehe durchaus, dass das ConciergePersonal für den Erfolg deiner Geschäftsidee extrem wichtig ist. Dennoch vergisst du auch, dass die Bootsanbieter und die nachfrager ebenfalls von grosser Bedeutung sind. Befinden sich auf der Plattform keine Bootsanbieter so ergibt sich keinen Nutzen für Bootsnachfrager und umgekehrt, wodurch sich die Plattform niemals etablieren wird. Ich denke, dass du das ebenfalls in deiner Lösung einarbeiten solltest. Die Massnahmen empfinde ich als eher verwirrend und ergeben in Zusammenhang mit vorherigen Kapiteln nicht wirklich Sinn. Du solltest hierbei ein wenig genauer werden und darauf achten, dass die Massnahmen gerade Punkt 1 und 3 mit bereits Beschriebenem übereinstimmt.

- *The student illustrates emotions throughout the entire review and uses strong words to demonstrate his emotion (e. g. concern) ("du musst zwingend eine Tabelle erstellen" "du solltest", "eher verwirrend", "ergeben nicht wirklich Sinn", "niemals", etc.)*
- *The student is using a lot of personal expressions ("Ich verstehen durchaus", "Dennoch vergisst du auch", "Ich denke, dass du", "empfinde ich")*

The dataset will be annotated according to the above-mentioned empathy elements on each *component* of the student-written peer review. This means that the evaluation of empathy will be applied to the description of the strengths, weaknesses and suggestions for improvement and will not be applied on a word or sentence basis. The following chapter will provide more detail on how to annotate the peer reviews.

5. Annotation process

The previous sections briefly described the components that are aimed to annotate in this study. For annotating these components, the annotation process is split into three steps: 1) reading of the entire review 2) labeling of the components and elaborations and 3) classification of both empathy scales.

1. *Reading of the entire peer review:* The annotators are confronted with the student-written peer review and are asked to read the whole document. This helps to get a first impression of the review and to get an overview of the single components and structure of it.

2. *Labeling the components and elaborations:* After reading the entire student-written peer review, the annotator is asked to label the three different components (strengths, weaknesses and suggestions for improvement). Details on how to label them can be found in the next chapter. Every supporting sentence (such as explanation, example, etc.) will be annotated together with the according component. Figure 2 illustrates how the components are annotated.

3. *Classification of both empathy scales:* Each component is assessed on its level of cognitive and emotional empathy by giving a number between 1-5. Each category is carefully defined and delimited.

Figure 2: Illustration of the annotation process (own illustration)

6. Review component boundaries

The structure of the student-written peer reviews can vary. Students were given the task to evaluate the peer's activity according to its strength, weaknesses and to give suggestions for improvement. They were not told on how to structure their reviews or on how long it should be. Therefore, significant differences can be detected in terms of structuring and length. Nevertheless, some guidelines can be derived to enhance annotation results. The following illustration helps to better understand them.

Figure 3: Illustration of the boarders of review components (own illustration)

General

- Generally, the review will be annotated according to three components (1: Strength, 2: Weakness, 3: Suggestions for improvement). Thus, every sentence in the first chapter will be annotated as a strength, every sentence in the second chapter as a weakness and every sentence in the third chapter as a suggestion for improvement. See case 1 in figure 3 as an example. However, there are two exceptions:
  - o Sentences that clearly show a strength but are mentioned in the weakness or suggestions for improvement component. Those exceptions should be annotated according to their allocation. See case 3 in figure 3 as an example ("Otherwise, all great!")
  - o If weaknesses and suggestions for improvement are both combined in one component (e. g. Schwächen & Verbesserungsvorschläge), the whole component will be annotated as a weakness. See case 2 in figure 3 as an example.
  - o Some reviews might not follow a clear chapter structure. Those reviews are annotated to the best of the annotator's knowledge according to the rules applying to strength, weakness and suggestion for improvement. See case 2 in figure 3 for an example, where the review is not structured

according to the three chapters. The guidelines respecting the distinction between strength, weakness, suggestion can be found below. These rules only apply to reviews that are not following a clear structure.

- Greetings, names or farewell sentences (such as "Dear XX", "best wishes", "kind regards", etc.) will not be annotated. See figure 2, where "good luck" is not annotated.

- Titles of the chapters (such as "Strength", "Suggestions for improvement"), numerations ("1.", "2.", "3."), but also other titles from the task (such as "Positive", "Key Resources", "Personas", "BPMN", etc.) will be ignored. See figure 2, where "1. Worin sehen Sie Stärken?" or case 2 in figure 3, where "Datenmodell" and "Funktionale Anforderungen" are not annotated

- Quotation marks at the beginning and at the end of the review will not be annotated. See figure 3, where " at the beginning and " at the end are not annotated. Also, any other marks ("→" ">") will be ignored.

- Further information that do not belong to the review from a contextual point of view (such as "First, I'd like to let you know that I have previous knowledge in the field of your business idea", "This leads me to the following strengths" "I have the following suggestions for you") will not be annotated. See case 2 in figure 3 where "Jedoch habe ich ein paar Anmerkungen" is not annotated.

- A sentence can consist of several components. The annotator is allowed to separate the sentence according to the component (e. g. "Ich finde deine Idee sehr gut, aber mir gefällt deine Darstellung nicht" is split into a strength component ("Ich finde deine Idee sehr gut") as well as a weakness component ("aber mir gefällt deine Darstellung nicht").
  - o A component can consist of several sentences. These sentences can be directly followed by each other or can be separated trough other components.
  - o Explanations, further elaborations, details, examples, etc. that support a component are annotated together with the component.

- Each component is assessed on its level of cognitive and emotional empathy
  - o When annotating, it is important that the entire component (every sentence that has been marked as this component) will be evaluated. Components are not split up for the assessment of empathy. See figure 2 and 3 as examples, where the sentences marked in green are all combined and together evaluated on their scale of emotional and cognitive empathy (and therefore only given one total score for each label per component).
  - o If one of the components is missing (e.g. the peer review does not include any strengths), it will not be annotated and therefore not given any label regarding emotional and cognitive empathy.

Strengths

- Something positive about the peer's submitted work ("Your BPMN is very structured", "Your idea is very interesting", etc.)
- Something that the peer liked (e. g. "I like how you did")

- Can be general or very specific
- A "positive" weakness or suggestion for improvement (e.g. "I do not find any weakness")

Weakness2

- Something negative about the peer's submitted work, a point of criticism ("I do not see why", "I do not understand", "It does not make sense")
- Missing parts or thoughts in the peer's idea or task, something that the reviewee wish he would have done or further hypothetical considerations ("I would have liked that you", "I wish you would have", "I missed", "It would be interesting to see/to know", "one could integrate", I think it would be good to")
- Questions that are showing disagreement ("Wouldn't it be?", "Don't you think that?")
- No concrete instruction, order or action for the peer derived yet, *no personal form of address* ("you", "the author")

Suggestions for improvement

- Something that should be added for the second version of the peer's work ("For the second version", "You should add", etc.)
- Concrete suggestions or parts that should be improved or need more attention ("I suggest that *you*", "my suggestion is that *you*", "try to", I think it would be good that *you*")
- Concrete instructions, invitations, orders or actions, directed towards the peer ("*You* could", *You* should", "*The author* must", "Would it be possible that *you*", the use of direct instructions like "add a second box" or "integrate another sentence about", etc.)

---

2 A weakness often implies a suggestion, too. If is more of a point of criticism rather than a concrete instruction or suggestion directed to the peer, the sentence stays a weakness. A suggestion must be explicitly directed towards the peer ("you", "the author").

7. References

Baron-Cohen, S., and Wheelwright, S. 2004. "The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences," *Journal of Autism and Developmental Disorders* (Vol. 34).

vom Brocke, J., Maaß, W., Buxmann, P., Maedche, A., Leimeister, J. M., and Pecht, G. 2018. "Future Work and Enterprise Systems," *Business and Information Systems Engineering* (60:4), pp. 357–366. (https://doi.org/10.1007/s12599-018-0544-2).

Davis, M. H. 1983. "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach.," *Journal of Personality and Social Psychology* (44:1), pp. 113–126. (https://doi.org/10.1037//0022-3514.44.1.113).

Decety, J., and Jackson, P. L. 2004. "The Functional Architecture of Human Empathy.," *Behavioral and Cognitive Neuroscience Reviews*, pp. 71–100. (https://doi.org/10.1177/1534582304267187).

Fadel, C., Bialik, M., and Trilling, B. 2015. *Four-Dimensional Education : The Competencies Learners Need to Succeed*.

Konrath, S. H., O'Brien, E. H., and Hsing, C. 2011. "Changes in Dispositional Empathy in American College Students over Time: A Meta-Analysis," *Personality and Social Psychology Review* (15:2), pp. 180–198. (https://doi.org/10.1177/1088868310377395).

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., and David, A. S. 2004. "Measuring Empathy: Reliability and Validity of the Empathy Quotient," *Psychological Medicine* (34:5), pp. 911–919. (https://doi.org/10.1017/S0033291703001624).

OECD. 2018. *The Future of Education and Skills - Education 2030*. (https://doi.org/2018-06-15).

Peterson, R. T., and Limbu, Y. 2009. "The Convergence of Mirroring and Empathy: Communications Training in Business-to-Business Personal Selling Persuasion Efforts," *Journal of Business-to-Business Marketing* (16:3), pp. 193–219. (https://doi.org/10.1080/10517120802484551).

Spreng, R. N., McKinnon, M. C., Mar, R. A., and Levine, B. 2009. "The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures," *Journal of Personality Assessment* (91:1), Taylor and Francis Inc., pp. 62–71. (https://doi.org/10.1080/00223890802484381).