

# Adapting Performance And Emotional Support Feedback To Cultural Differences

Muhammad Adamu Sidi-Ali<sup>1</sup>, Judith Masthoff<sup>2</sup>, Matt Dennis<sup>3</sup>, Jacek Kopecky<sup>3</sup>, and Nigel Beacham<sup>1</sup>

<sup>1</sup>University of Aberdeen (UK)

<sup>2</sup>University of Aberdeen (UK) and Utrecht University (NL)

<sup>3</sup>University of Portsmouth (UK)

## Abstract

This paper investigates adaptation of feedback to learners' cultural backgrounds. First, we investigate how to portray the cultural background of a learner. Second, we present a qualitative focus-group study, investigating how participants from different cultures believe culture affects the kind of feedback given to a learner. Finally, we present an empirical study on how humans adapt feedback based on the cultural background of learners to inspire an algorithm. Our investigations resulted in a set of stories which can be used to reliably portray a person's culture when investigating cultural adaptation in indirect experiments and user as wizard studies. They also provided insights into the adaptations people make to cultural differences.

## 1 Introduction

Online teaching and learning is popular today. As more learners participate in online learning, there is a need to keep them motivated, as learner retention is a challenge for this medium [25]. One way of keeping learners motivated is by providing adapted feedback on performance, since feedback is an important part of learning and motivation [24].

Feedback can be adapted to different learner characteristics, including performance, affective state, personality, and so on [9]. In this paper, we are particularly interested in adapting feedback to learners' cultural backgrounds. Cultural background has been shown to be important in the acceptance of user interfaces [17, 10], information systems [2], and e-learning systems [1]. Researchers have highlighted that learners' cultural backgrounds affect their behaviours when using educational systems. For example, studies have shown that students with different cultural backgrounds have different help-seeking behaviours [26] and off-task behaviours [30] when using intelligent tutoring systems. Another study by Kim and Bonk [19] revealed some cross-cultural differences in online collaboration of participants such as being social, group focused, and being pragmatic and action oriented. Pratt et al. [29] reported that knowledge, teaching and learning are deeply rooted in cultural backgrounds and social structures of learners. Therefore, in educational psychology, culture has always been acknowledged as a significant factor in learning and development [37].

Where the main objective of a computer-based information system is to improve human performance [23], the main aim of adaptation in e-learning is to increase performance, motivation, and knowledge in general [18]. In this paper, we describe focus groups and User-as-Wizard studies which investigate how humans think feedback should be adapted to learner cultural background and performance, specifically Power Distance, Individualism and Uncertainty Avoidance from Hofstede [14]. We convey the cultural background of the learner using stories, the development of which is described in section 3. Section 4 describes a series of focus groups which asked participants, themselves from a variety of cultural backgrounds, to provide performance feedback and emotional support (from options developed in previous research [8]) to the learners conveyed by the stories, at a variety of performance levels. Following on from these studies section 5 describes a set of User-as-Wizard [22] studies where 150 participants playing the role of a teacher gave feedback to learners from different cultural backgrounds, who had 'just passed' a test. We hope to use the insights gained from these studies to develop an algorithm for use in a future ITS so that it can utilize these adaptations and improve learner motivation.

## 2 Related Work

E-learning systems have been adapted to various learner characteristics. For example, Dennis et al. [8] adapted feedback to learner personality and performance to increase motivation. Their work investigated how personalised feedback on performance may be given to a learner by a conversational agent, and how learner motivation can be maintained by incorporating the most effective emotional support statements in the feedback. Okpo et al. [28] investigated the adaptation of exercise selection to learner self-esteem, effort, and performance. Grawemeyer et al. [12] developed an affective state reasoner to adapt different types of feedback to a learner's affective state during their interaction with a learning environment.

Gouli et al. [11] proposed an adaptive feedback framework (AFF) that provides feedback using the knowledge level, preferences and interaction behaviour of learners as main sources of adaptation. Vasilyeva et al. [36] outlined learner characteristics that are important for individual feedback adaptation in e-learning systems as: personal data, knowledge, and interaction parameters. Personal data such as age, gender, etc, for the purpose of giving the application a personal touch such as including a learner's name in the feedback, knowledge referring to what the learner already knows about the topic or related topics, and interaction parameters as time spent viewing pages with learning materials, total time spent on the assignment or number of attempts to pass the test or assignment.

Okpo et al. [27] investigated characteristics that may possibly be considered when selecting exercises for learners and how humans adapt exercise selection to learner personality and performance, to enable an Intelligent Tutoring System (ITS) to tailor exercise difficulty to these characteristics.

However, despite all this work on adaptation to learner characteristics, there is no real body of work on adaptation to cultural differences. Social scientists including Hall [13], Trompenaars and Hampden-Turner [35], Hofstede and Hofstede [15] have provided models for measuring culture. Hall classified culture into low and high context based on the predominant style of communication in a society. Messages in low context societies are spelled out and defined, while in high context societies, message cannot be understood without a great deal of background information. He classified Western societies as low context, and Asian cultures as high. However, there is no large-scale cross cultural study that has used this model to classify existing cultures into low and high context. The model introduced by Trompenaars and Hampden-Turner, like many other models, has been reviewed by Taras et al. [33], and was found to be conceptually similar to the dimensions introduced by Hofstede.

Hofstede provided a bipolar model of cultural dimensions (CDs) that have been consistently and successfully used to assess cultural differences. There are many differences and similarities in cultures of people, after a review of literature, it is clear that the CDs of Hofstede cover most aspects [4, 32]. Over 100,000 employees of IBM in 66 countries participated in the cultural dimensions survey of Hofstede. He used the scores of the survey to classify each country relative to others on each dimension. The dimensions are: Power distance, Individualism, Masculinity, Uncertainty avoidance, and Long-term orientation[15].

**Power Distance (PD)** The degree to which less powerful people in institutions and organizations in a country expect and accept the unequal distribution of power. Countries that score low (small power distance) have minimized inequality amongst its people, while countries that score high (large power distance) expect inequality amongst its people.

**Individualism (IND)** Defined as the extent to which members of a society maintain interdependence. People in countries that score low (Collectivist) think in terms of "we", while People in countries that score high (individualist) think in terms of "I",

**Masculinity (MASC)** The extent to which a society is driven by competition, achievement and success, with success meaning the best in a field or a winner. Masculinity does not mean male, and femininity female.

**Uncertainty Avoidance (UA)** Defined as the extent to which members of a culture feel threatened by confusing or unknown circumstances and hold-on to beliefs to avoid these. Countries that score low (weak uncertainty avoidance) are comfortable in ambiguous situations and unfamiliar risks, while on the other side, countries that score high (strong uncertainty avoidance) fear ambiguous situations and unfamiliar risks.

**Long-Term Orientation (LTO)** The extent to which a society maintains links with its past when dealing with its present and future challenges. Societies that score low (short-term Orientation) have great respect for tradition and are hesitant about societal change, while those that score high (Long-Term Orientation) have respect for circumstances and encourage having modern education to prepare for the future.

### 3 Story Creation and Validation

To conduct our research on the impact of culture on feedback, we needed a way to portray culture. We decided to produce stories for culture in a way similar to the personality stories produced by Dennis et al. [6].

#### 3.1 Story Creation

The stories were created from the 26-item Individual Cultural Values Scale (CVSCALE) [38] by putting phrases and sentences together. While creating the stories, the interested cultural dimension and whether the story was for a low or high level of the dimension were taken into consideration. Common male names were added to each story to make it more real. Examples of the created stories are shown in Table 1.

#### 3.2 Validation Study

Following story creation, a validation study was performed to ensure that the correct dimension was conveyed at the correct level for each story.

##### 3.2.1 Study Design

A mixed design was used – one group of participants saw all the low stories, and the other saw all the high stories<sup>1</sup>. Within these groups, each participant saw five stories, one from each cultural dimension. Participants read the story and indicated how well the dimension was portrayed using a scale.

##### 3.2.2 Participants

A convenience sample of participants was recruited through email and social media. There were 56 participants for round 1: 32 for the low stories and 23 for the high. In the second round, there were 40 participants: 20 for low stories and 20 for high.

##### 3.2.3 Variables

The independent variable in this study is the Story: one for each dimension. The dependent variable in this study is the extent to which participants feel the fictional person in the story belongs to either a low or a high dimension, lowest possible vs. highest 5.

##### 3.2.4 Hypotheses

- H1 There will be a significant difference in the perceived level of the target dimension between the low and the high stories
- H2 The mean for the low stories will be at the low end of the scale (under 3)
- H3 The mean for the high stories will be at the high end of the scale (above 3)

---

<sup>1</sup>This means that for each dimension a between subject design was used, in line with the method for story validation in [31]

Table 1: Cultural dimensions stories. \* indicates an adjusted story following R1 validation

CD	Story
Low PD	In the society John grew up in, he was taught that his opinion was as important as his parents' and not to take things for granted. People in this society must create their own functions at work, must not take their bosses' decisions for granted, and always question their bosses' actions. In addition, it is widely accepted that the most effective way to change a political system is through public debates and free elections.
High PD	In the society John grew up in, he was taught never to question his parents' authority and to accept the authority of older or important people. In this society, people have clearly defined roles at work, the boss makes all decisions, and everybody accepts and respects the boss. In addition, people in this society believe that the most effective way to change a political system is to replace those in power through drastic means.
Low IND	In the society Alex grew up in, the behaviours of people are influenced by the strong loyalty, conventions, and rules of the groups they belong to. People are promoted and recognized based on their loyalty and age, and it is immoral for a boss not to offer a job to a relative. Generally, people are concerned with what others think about them.
High IND	In the society Alex grew up in, everybody has full personal freedom and people choose their friends based on common likes/dislikes/interests. People are concerned only with their own rules and objectives. Generally, people are promoted based on competence no matter their age, and it is immoral for a boss to offer a job to a relative.
Low MASC	Jordan grew up in a society where a good quality of life is important for both men and women, and people seek love and mutual affection from their partners. At work, it is a must for decisions to be based on consensus, and people are motivated by a relaxed and friendly atmosphere. People who do not win are sympathised with, and others who succeed are envied.
High MASC	Jordan grew up in a society where men focus on material success while women are concerned with the well-being of others. In relationship, what people want the most from their partners is support in difficult situations. At work, people have clear objectives and evaluation systems for what they accomplish, and conflict is seen as positive and productive.
Low UA*	Sam grew up in a society where there are few rules. In his society, people DO NOT plan (the future can never be known do not try to control it just let it happen), and are welcoming to whatever happens. When going out, people DO NOT have to carry identity cards. People are relaxed and DO NOT normally express emotions in public. In addition, the society appreciates people who adapt to different circumstances.
High UA*	Sam grew up in a society where there are many rules and customs people must respect. He was taught to be organised and be in control, as people always want to know exactly what to expect. In his society, differences make people highly anxious, and it is OK to show emotions in public. When going out, people have to carry identity cards, and the society appreciates high competence and expert leadership.
Low LTO*	Peter grew up in a society where people always want consistency in the information they are presented with. In this society, people believe in quick rewards. Their behaviours are always influenced by their culture, causing them to clearly know what is good or bad. In addition, they value personal stability and continuity.
High LTO*	Peter grew up in a society where children are taught to ask WHAT and HOW. In this society, people work on common goals with no concern for what is good or bad, and can live with contradictory information. They work for future rewards, and are comfortable with slow results from continuous efforts. In addition, they think everything is relative and permanently changing.

Table 2: Perceived dimension levels for the stories. *t*-test significance values between low and high stories, Bold indicates significance ( $p < 0.001$ ) between mean scores. \* indicates modified stories validated in round 2.

Dimension	Story level	<i>N</i>	mean (StDev)
Power distance	Low	32	<b>2.5</b> (0.9)
	High	23	<b>4.0</b> (0.7)
Individualism	Low	32	<b>2.2</b> (0.7)
	High	23	<b>3.6</b> (0.7)
Masculinity	Low	32	<b>1.8</b> (0.7)
	High	23	<b>3.5</b> (1.0)
Uncertainty avoidance*	Low	20	<b>2.0</b> (0.5)
	High	20	<b>3.8</b> (0.7)
Long-term orientation*	Low	20	<b>1.8</b> (0.4)
	High	20	<b>3.8</b> (0.9)

### 3.2.5 Procedure

Participants read a story about a fictional person’s culture. Next, they filled in a set of five point Likert scales with polarized statements on both ends of the scale<sup>2</sup>. An example of a pair of statements for the PD dimension is: “Parents treat children as equals” and “Parents teach children obedience”. For each scale, they indicated the extent they felt the society of the person depicted in the story is more like the one on the left or the right. The statements used came from the key differences of societies by Hofstede et al. [16]. We used 7 or 8 statement pairs per dimension.

### 3.2.6 Results

We conducted two-sample *t*-tests and the results from three out of the five pairs of stories (PD, IND, MASC) were significantly different in the perceived level of the target dimensions between the high and the low stories ( $p < 0.001$ ) and the means were low enough for the low stories and high enough for the high stories so all hypotheses are accepted (see Table 2).

For UA and LTO, the stories failed to validate. These stories were modified and re-validated in a round 2 of the study. For example, in the low UA story, “the society appreciates those that could move in different environments” was replaced with “the society appreciates people who adapt to different circumstances”. These changes were to eliminate sentences that participants found vague and to make the story express the whole concept of the target cultural dimension, these corrected stories are marked with \* in Table 1. Following this, these stories successfully validated as shown in Table 2.

## 4 Focus Groups

This study investigates how feedback is given by people from different cultures to hypothetical learners exhibiting different cultural dimensions. Focus groups were used for this investigation because they are known to produce a well-organized and appropriate insight on a topic [20, 21]. We based the focus group (FG) composition on Hofstede’s cultural dimensions. We conducted eleven FGs to compare three dimensions: Power Distance, Individualism and Uncertainty Avoidance. These dimensions were chosen given our access to participants from countries that were either high or low on these dimensions. Table 3 shows the composition of the FGs and the dimension investigated in them.

### 4.1 Participants

There were 38 participants across all the focus groups. For each FG, we recruited students who according to Hofstede’s work came from cultures that are similar on a particular dimension. We conducted at least one FG high on the

<sup>2</sup>So a kind of a semantic differential scale

Table 3: Composition of Focus Groups

Group	#	♂	♀	Country	Dimension
FG-NL1	2	2		Netherlands	PD
FG-NL3	2		2	Netherlands	PD
FG-CH1	4	4		China	PD
FG-NN	4	4		Nigeria (N)	PD
FG-SWN	3	2	1	Nigeria (SW)	PD
FG-INDO	5	3	2	Indonesia	IND
FG-UK	4	3	1	United Kingdom	IND
FG-NL2	3	2	1	Netherlands	IND
FG-CH2	2	1	1	China	UA
FG-SA	2		2	Saudi Arabia	UA
FG-NL4	3	2	1	Netherlands	UA
FG-F&S	4	3	1	France & Spain	UA

Table 4: Performance feedback from Dennis et al. [5].

Your performance is	substantially above	my expectations
	above	
	slightly above	
	meeting	
	slightly below	
	below	
	substantially below	

dimension and one low on the dimension. Each FG contained from 2 to 5 persons and lasted about 45 minutes. See Table 3 for demographics. Ethical approval was obtained from the physical sciences and engineering ethics board of the University of Aberdeen.

## 4.2 Research Question

Do people adapt feedback to cultural dimensions? And if so, in which way?

## 4.3 Materials

The following materials were used in conducting the focus groups:

**Cultural stories** The validated stories were used to make participants aware of the cultural backgrounds of fictional learners, to whom feedback will be given. (See Table 1)

**Performance cards.** In a test with 50% as pass mark, participants were asked to give the fictional learner feedback for scoring 10%, 45%, 55%, and 95%.

**List of performance and emotional support (ES) statements** Participants agreed on the most appropriate feedback to give the learners from a list of performance [5] and emotional support [7] statements, see Tables 4 and 5. They were not informed which category each emotional support statement belonged to, and the statement list was randomized.

## 4.4 Procedure

Participants were told that the FG's purpose was to investigate whether a learner's cultural background should affect the feedback given to that learner. Participants were given the stories of two fictional learners, one belonging to a low

Table 5: Emotional support feedback from Dennis et al. [7].

Praise	P1	Well done
	P2	That was hard but you did it
	P3	I am proud of you
Reassurance	R1	You will get the hang of it eventually
	R2	Everyone is wrong sometimes
	R3	Everyone finds this hard
Emotional Reflection	ER1	I know what you're feeling
	ER2	I understand that you may be upset
	ER3	You must be really happy
Advice	A1	Just read the questions more carefully
	A2	Just take a bit longer next time
	A3	Just keep practicing

and the other belonging to a high dimensional society on the same dimension (see Table 3). In each FG, participants decided which of the fictional learners' cultures is more like theirs. Next, they were shown the performance of that learner, and asked to agree on appropriate feedback from the list of performance and emotional support feedback provided. They were asked to select one performance feedback statement and add as many emotional support statements as they wished.

## 4.5 Results

### 4.5.1 Power Distance

Table 6 shows the comparison between the Dutch, Chinese, northern and southwestern Nigerian FGs. The Netherlands has a low PD culture, while China and Nigeria are both high PD cultures.

With the exception of FG-NN, all FGs felt PD mattered, and provided different feedback for the high PD learner than the low PD learner. Except for FG-NL3, they did this for all performance levels. FG-CH1 provided more Advice for all scores to the high PD learner than the low PD learner who received more Emotional Reflection or Praise. FG-SWN also provided more Advice to the high PD learner than the low PD learner who received more Emotional Reflection or Reassurance (except for the 95% score). FG-NL3 also provided more Advice for the badly failed score to the high PD learner, again provided more Reassurance to the low PD learner. In contrast, FG-NL1 provided more Advice to the low PD learner. FG-NL3 used praise for both passing scores, but adapted the praise statement used, opting for more enthusiastic statements for the low PD learner. FG-CH1, FG-SWN and FG-NL1 were all harsher in their performance feedback for the passing scores for the high PD than the low PD learner. FG-SWN was also harsher for the failing scores. FG-CH1 explained this, stating that for example using "substantially above" at 95% is not proper, as 95% is an average score in China. One of the participants even stated that he has never heard his teacher say "your performance is above my expectation". FG-NL1 said that they felt harsher, less positive feedback needed to be given to the high PD learner as that culture was more authoritarian.

### 4.5.2 Individualism:

Table 7 shows the comparison between the Indonesian, UK and Dutch FGs. Indonesia has a low IND (so, collectivist) culture, while the UK and the Netherlands both have a high IND culture.

Two of the three FGs clearly took IND into account; the other (FG-UK) only made slight changes. FG-INDO provided more emotional support (in particular emotional reflection, reassurance and praise) for all scores for the low IND than for the high IND learner, with a very high number of statements for failing learners. FG-NL2 also provided more emotional support for the low IND learner, but only for the badly failed and did well scores. More than one participant of FG-INDO repeatedly said "We are emotional people", and the rest of the group members agreed. FG-INDO was also more positive in their performance feedback for the low IND learner for two scores. For the 10% score, FG-INDO agreed on 'your performance is below my expectation' because they felt 'substantially below' would be too hard emotionally for a learner from a low IND culture. In contrast, the other two FGs (which had a high IND

Table 6: Power Distance Focus Group Results

country	score	level	performance	emotional support
Netherlands ( <i>low</i> )	10%	low high	substantially below	ER2, A3 ER2
FG-NL1	45%	low high	slightly below below	A2, A1 R1
	55%	low high	meeting slightly below	P1 P2
	95%	low high	substantially above above	P3 P1
Netherlands ( <i>low</i> )	10%	low high	substantially below	R1 A3
FG-NL3	45%	both	slightly below	A2
	55%	low high	slightly above	A1 A3
	95%	both	above	ER3
China ( <i>high</i> )	10%	low high	below	ER2, A3 A1, A3
FG-CH1	45%	low high	below	ER2, A1 A2, A3
	55%	low high	slightly above below	P2, A1 A3, A1
	95%	low high	substantially above meeting	ER3, P3, P1 A3, P3, P1
Nigeria (N) ( <i>high</i> )	10%	both	substantially below	ER1, ER2, R2, A3
	45%	both	slightly below	R1, A2, A1
FG-NN	55%	both	meeting	P2, A3, P1
	95%	both	substantially above	ER3, P3, P1
Nigeria (SW) ( <i>high</i> )	10%	low high	below substantially below	ER2, R3, R2, A1 ER2, A1, A2, A3, R1
FG-SWN	45%	low high	slightly below below	A2, A3 ER2, A1, A2, A3, R1
	55%	low high	meeting slightly below	R2, A3, R1 A3, A2, A1
	95%	low high	substantially above meeting	ER3, P3, P1 ER3, P2, P3, P1



Table 7: Individualism Focus Group Results

Country	score	level	performance	emotional support
Indonesia ( <i>low</i> )	10%	low high	below substantially below	ER1, ER2, R3, A3 None
FG-INDO	45%	low high	slightly below	ER1, ER2, R3, A2, A1 A1, A3
	55%	low high	meeting slightly below	P2, P1 P1
	95%	low high	substantially above	P2, ER3, P3, P1 ER3, P3, P1
UK ( <i>high</i> )	10%	low high	substantially below	A3, ER2 ER2, A3
FG-UK	45%	low high	slightly below	R1, A2 R1, A3
	55%	low high	meeting	P2, A3 P2, A2
	95%	both	substantially above	ER3, P1
Netherlands ( <i>high</i> )	10%	low high	substantially below	P3, A3 R3
FG-NL2	45%	low high	slightly below	R1 A3
	55%	low high	slightly below meeting	A3 P1
	95%	low high	meeting above	P3, ER3 P3

Table 8: Results from Uncertainty Avoidance Focus Groups

Country	score	level	performance	emotional support
China ( <i>weak</i> )	10%	weak strong	substantially below	A3, A1 ER2, R2, A3
FG-CH2	45%	weak strong	slightly below	A3, A1, ER2 ER1, R1, A2
	55%	weak strong	slightly below below	A3, A2, A1 A3, A1
	95%	weak strong	substantially above meeting	ER2, A3, A1 A3, P1
Saudi Arabia ( <i>strong</i> )	10%	weak strong	substantially below	R2 ER1, A3
FG-SA	45%	weak strong	slightly below below	A1, A3 ER2, A1, A2
	55%	weak strong	meeting slightly below	P2 R3, A2, A3
	95%	weak strong	above meeting	P1 P2, ER3, P3, P1
Netherlands ( <i>neutral</i> )	10%	weak strong	substantially below slightly below	A1, A3, R1 ER1, A3
FG-NL4	45%	weak strong	slightly below	R2 R3, A2
	55%	weak strong	meeting	P2, P1 ER3, P1
	95%	weak strong	substantially above above	P3, ER3 P3
France & Spain ( <i>strong</i> )	10%	both	substantially below	A3
FG-F&S	45%	both	below	ER2, R1, A3
	55%	both	meeting	P2, P1
	95%	both	substantially above	ER3, P3, P1

culture) had no problem with using ‘substantially below’, and never used more than two emotional support statements. In contrast to FG-INDO, FG-NL2 were more positive in their performance feedback to the high IND learner, though only for the higher scores. FG-NL2 explained their additional use of emotional support saying that the low IND learner would care more about what others were thinking about them.

#### 4.5.3 Uncertainty Avoidance:

Table 8 shows the comparison between the Chinese, Saudi Arabian, Dutch and French/Spanish FGs. China has a weak Uncertainty Avoidance culture, the Netherlands is neutral, while Saudi Arabia, France and Spain have a strong Uncertainty Avoidance culture.

With the exception of FG-F&S, all FGs took UA into account, and for all scores. FG-CH2 provided more advice for all scores to the weak UA learner than the strong UA learner who received more emotional reflection, reassurance and praise. They stated that the strong UA learner seems to obey rules therefore needs more leniency and sympathy. FG-NL4 also provided more advice for the strong UA learner, but only for a just failing score. FG-SA gave more emotional support to the strong UA learner than the weak UA learner, stating that the weak UA learner adapts to any

circumstance, so needs less support. FG-NL4 argued that the strong UA learner is rather sensitive compared to the low UA who can handle criticism better. They made the performance feedback less harsh for the strong UA learner for the badly failing score, and also provided emotional reflection rather than advice in that case.

## 5 User-as-Wizard Study

The FGs provided evidence that people feel feedback should be adapted to the learner’s cultural background (PD, UA, and IND) and provided suggestions on the kinds of adaptations that would be required. However, FGs by nature only contain a limited number of participants, and we wanted to investigate whether these results could be replicated in a larger scale empirical study. Ultimately, we would like to run studies in different countries with participants who differ on these cultural dimensions. This was not feasible yet, so we started with participants of one particular country, namely the US, which is very high on IND, low on PD, and just below the medium on UA. We focus this study on one performance level, namely 55% (just passed), as our previous research on personality traits [8] showed that most adaptations are found for scores close to the pass-mark.

### 5.1 Methodology

For this study, we utilize the User-as-Wizard approach [22] where participants take the role of a system providing feedback to learners. We utilized a between-subjects design— each participant saw one story about a learner and their performance on a class test. Similar to the studies outlined in section 4, participants had to provide one performance feedback statement and could add emotional support statements if they wished.

In this study, we investigated the same three cultural dimensions as in section 4: Power Distance, Individualism and Uncertainty Avoidance at high and low levels, thus there were 3x2 conditions.

### 5.2 Participants

150 participants took part in this study with 25 per condition: 44% identified as female, 55% male and 1% as other. 13% were aged 18-25, 63% 26-40, 23% 41-65 and 1% over 65. 7% of participants were teachers, and 10% were students. Participants were crowd-sourced through Amazon Mechanical Turk and paid \$0.50. To be eligible for the study, participants had to be based in the United States, have an acceptance rate of 90%<sup>3</sup> and pass a cloze-test [34] for English fluency, due to the language based nature of the study. The average time taken to complete the study was 3 minutes and 16 seconds<sup>4</sup>.

### 5.3 Materials

The dimension (PD, IND or UA) and level (low or high<sup>5</sup>) were conveyed using the corresponding story from Table 1. Performance Feedback was formed using the options shown in Table 4. As in section 4, emotional support could be added using any number of statements from Table 5.

### 5.4 Procedure

Participants began by completing a consent form and the English Test. Those who passed were then directed to the study proper. Participants were shown one story about the learner’s cultural dimension and level, and told that the learner had scored 55% on a test (with a 50% pass mark). Participants were asked to give one performance feedback statement. Subsequently, participants could add emotional support statements if they wished. The study finished by collecting basic demographic information and any other comments from participants.

<sup>3</sup>meaning that 90% of the tasks they do are judged as acceptable quality

<sup>4</sup>excluding time taken to complete the English Test

<sup>5</sup>The levels for Uncertainty Avoidance are *weak* or *strong* but are referred to as low or high respectively in this section for clarity and consistency

Table 9: Results for Adaptation Study. Bold CD = significant overall effect ( $p < 0.05$ ) for level from  $\chi^2$  for performance feedback. \* = adjusted residual  $\pm 2.4$ . † = significant effect of level on Emotional Support from Mann-Whitney U Test,  $p < 0.03$

Key: Su = substantially, Sl = slightly, M = Meeting, B = Below, A = Above.

CD	level	performance (% participants)							mean ES statements used (SD)				
		Su.B	B	Sl.B	M	Sl.A	A	Su.A	numA	numER	numP	numR	numS
PD	low	0	16	28	20	24	8	4	0.76 (0.66)	0.04 (0.20)	0.52 (0.59)	0.24 (0.44)	1.56† (1.00)
PD	high	4	24	12	16	40	4	0	0.96 (0.61)	0.12 (0.44)	0.80 (0.76)	0.40 (0.58)	2.28† (1.14)
IND	low	4	12	8	32	40	4	–	0.84 (1.03)	0.08 (0.23)	0.92 (1.04)	0.12 (0.33)	1.96 (0.98)
IND	high	4	12	16	48	12	8	–	1.12 (0.88)	0.12 (0.33)	0.80 (0.87)	0.28 (0.46)	2.32 (1.18)
<b>UA</b>	low	–	8	0*	40	48	4	–	0.60 (0.76)	0.16 (0.37)	1.04† (0.89)	0.36 (0.64)	2.16 (1.41)
<b>UA</b>	high	–	24	20*	16	36	4	–	0.88 (0.67)	0.08 (0.28)	0.40† (0.58)	0.32 (0.48)	1.68 (0.90)

## 5.5 Variables

The independent variable was the level of the cultural dimension: low or high. The dependent variables were:

- Performance Feedback: The statement used to describe the learner’s performance (substantially below, below, slightly below, meeting, slightly above, above or substantially above expectations).
- The total number of each type of Emotional Support used: numA (Advice), numP (Praise), numER (Emotional Reflection), and numR (Reassurance).
- The total number of emotional support statements given overall (numS)

## 5.6 Hypotheses

The main hypotheses for this study are:

H1 The type of performance feedback given will differ based on the level of the cultural dimension of the learner

H2 The type of emotional support given will differ based on the level of the cultural dimension of the learner

H3 The total number of emotional support statements given will differ based on the level of the cultural dimension of the learner

In addition, based on the FGs’ results for the 55% performance level, we also hypothesize that:

H4 Harsher performance feedback will be given to the high UA learner than the low UA learner

H5 More Praise will be given to the low UA learner than the high UA learner

H6 Harsher performance feedback will be given to the high PD learner than the low PD learner

H7 More Advice will be given to the high PD learner than the low PD learner

## 5.7 Results

The results for the study are shown in Table 9. All  $p$  values are Benjamini-Hochberg corrected [3] with a false discovery rate of 25%.

Table 10: Amount of Emotional Support given for PD.

PD	numS (%participants)					
	0	1	2	3	4	5
low	4	56	28	8	0	4
high	0	28	36	20	12	4

Table 11: Amount of Praise given for UA.

UA	numP (%participants)			
	0	1	2	3
low	32	36	28	4
high	64	32	4	0

### 5.7.1 Performance Feedback

A  $\chi^2$  test for performance feedback  $\times$  level was performed for each of the three dimensions. The only dimension which showed significance ( $\chi^2(4, N = 50) = 10.00, p < 0.05$ ) was UA. As shown by Table 9, this appears to be caused by participants using more negative terms (behind or slightly behind expectations) for high UA. This confirms hypothesis H1 but only for UA, and also confirms H4. Examining the adjusted standardised residuals from the chi-square, 'slightly below' has a residual of  $\pm 2.4$  - with 20% of participants using this term in the high condition and none in the low condition. No other residuals are significant. No evidence was found for hypothesis H6.

### 5.7.2 Emotional Support

A Mann-Whitney U test for numA, numER, numP, numR and numS  $\times$  level was performed for each dimension. For Power Distance, numS was significant ( $U = 191, p < 0.03$ ) with those in the high condition giving more emotional support statements overall. From Table 10, most participants in the low condition gave one statement, whereas most participants in the high condition gave more than one statement. This supports hypothesis H3 but only for PD. For UA, numP was significant ( $U = 184.5, p < 0.02$ ), meaning the learner with low UA received more praise, as shown in Table 11, with most participants in the high condition giving no praise. This supports hypothesis H2 but only for UA, and confirms hypothesis H5. No support was found for hypothesis H7, though the trend in the number of Advice statements is in the expected direction.

## 6 Conclusions

This paper investigated whether cultural dimensions need to be taken into account when providing feedback to learners. Both our qualitative focus group study and quantitative User-as-Wizard study showed that people think this adaptation is needed. These studies also provided insights into the kinds of adaptations that may be required.

To perform these studies, we needed a way to depict cultural dimensions. Therefore, pairs of stories have been created and validated for each of Hofstede's five cultural dimensions, namely: Power distance, Individualism, Masculinity, Uncertainty avoidance, and Long term orientation. These stories can be reliably used to portray a person's culture when investigating cultural adaptation in, but not limited to, indirect experiments and User-as-Wizard studies. As has been shown in [31], these kinds of stories can also be used to obtain user characteristics, by putting low and high stories of a characteristic at each end of a slider scale. We intend to produce such a scale and measure its reliability using the method in [31]. This may provide learners with a simple and less intrusive way of indicating culture without the need to disclose a particular ethnic background for instance.

In the FG study, learners adapted feedback to all three cultural dimensions we investigated, namely PD, UA, and IND. Regarding PD, FGs tended to be harsher in performance feedback to high PD learners, and also provided more advice rather than other types of emotional support to such learners. Regarding IND, FGs tended to provide more

emotional support to low IND learners, particularly for low performance. Regarding UA, the adaptations depended a lot on performance level, and adaptations were made both to performance feedback and emotional support.

In the User-as-Wizard study, investigating a performance level of 55% (just passed), US based participants adapted both performance feedback and emotional support to PD and UA. No adaptation to IND was found, but this is not surprising, as the FGs' adaptations to IND were primarily for other performance levels. The adaptations made for UA were fully in line with the expectations based on the FGs, providing more praise to the low UA learner and harsher performance feedback to the high UA learner. For PD, we did not find the harsher performance feedback the FGs had provided for the high PD learner. Further studies will be required with participants from other cultures. We also need to repeat the study for the other performance levels, to see for example if IND does have an impact for failing learners as predicted by the FGs. Any adaptations discovered from empirical studies from the general population will need to be verified by teachers in a further studies before their incorporation into the resulting algorithm.

The main limitation of this work is that we so far have investigated what people believe would be good feedback for learners from different cultures. Of course, studies will also need to be conducted to directly measure the impact of feedback adaptations on learner performance and motivation. However, despite this limitation, the work presented is important as it is the first to address adaptation of feedback to culture.

The work presented in this paper also raises an ethical issue, as it may seem unethical to adapt feedback to culture. However, the FGs clearly indicated that people from certain cultures are more used to, and are expecting, a different kind of feedback, whether it is the more emotional feedback our Indonesian participants expected, or the harsher feedback expected by our Chinese participants. If future studies show that feedback adaptation to culture improves learner performance and motivation, this may alleviate some of the ethical concerns.

## Acknowledgements

The first author would like to thank the Petroleum Technology Development Fund (PTDF) of Nigeria for sponsoring his PhD.

## References

- [1] Aparicio, M., Bacao, F., and Oliveira, T. (2016). Cultural impacts on e-learning systems' success. *The Internet and Higher Education*, 31:58–70.
- [2] Baptista, G. and Oliveira, T. (2015). Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. *Computers in Human Behavior*, 50:418–430.
- [3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 57(1):289–300.
- [4] Clark, T. (1990). International marketing and national character: A review and proposal for an integrative theory. *The Journal of Marketing*, 54:66–79.
- [5] Dennis, M., Masthoff, J., and Mellish, C. (2012a). Adapting performance feedback to a learners conscientiousness. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 297–302, Montreal, Canada. Springer, Springer Heidelberg Dordrecht London New York.
- [6] Dennis, M., Masthoff, J., and Mellish, C. (2012b). The quest for validated personality trait stories. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 273–276, Lisbon, Portugal. ACM, ACM New York, NY, USA 2012.
- [7] Dennis, M., Masthoff, J., and Mellish, C. (2013). Does learner conscientiousness matter when generating emotional support in feedback? In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 209–214, Geneva, Switzerland. IEEE, IEEE.
- [8] Dennis, M., Masthoff, J., and Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, 26(3):877–931.

- [9] Dennis, M. G. (2014). *Adapting feedback to learner personality to increase motivation*. PhD thesis, University of Aberdeen.
- [10] Evers, V. and Day, D. (1997). The role of culture in interface acceptance. In *Human-Computer Interaction INTERACT'97*, pages 260–267, Sydney, Australia. Springer, Springer, Boston, MA.
- [11] Gouli, E., Gogoulou, A., Papanikolaou, K. A., and Grigoriadou, M. (2006). An adaptive feedback framework to support reflection, guiding and tutoring. In *Advances in web-based education: Personalized learning environments*, pages 178–202. IGI Global, Web.
- [12] Grawemeyer, B., Mavrikis, M., Holmes, W., and Gutierrez-Santos, S. (2015). Adapting feedback types according to students' affective states. In *International Conference on Artificial Intelligence in Education*, pages 586–590, Switzerland 2015. Springer, Springer, Cham.
- [13] Hall, E. (1976). *Beyond Culture, New York (Anchor Press) 1976*. Anchor Books, United States of America.
- [14] Hofstede, G. (1991). *Cultures and organizations. Intercultural cooperation and its importance for survival. Software of the mind*. McGraw-Hill, London.
- [15] Hofstede, G. and Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE Publications, Thousand Oaks California.
- [16] Hofstede, G., Hofstede, G., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, Third Edition*. McGraw-Hill Education, United States of America.
- [17] Ito, M. and Nakakoji, K. (1996). Impact of culture on user interface design. In *International users interface*, pages 105–126, New York, USA. John Wiley & Sons, Inc., John Wiley & Sons, Inc.
- [18] Kareal, M. and Klema, M. (2006). Adaptivity in e-learning. current developments in technology-assisted education, 260-264. retrieved april 9, 2013.
- [19] Kim, K.-J. and Bonk, C. J. (2002). Cross-cultural comparisons of online collaboration. *Journal of Computer-Mediated Communication*, 8(1):JCMC814.
- [20] Krueger, R. A. and Casey, M. A. (2002). Designing and conducting focus group interviews.
- [21] Masadeh, M. A. (2012). Focus group: Reviews and practices. *The Journal of Applied Science and Technology*, 2(10):63–68.
- [22] Masthoff, J. (2006). The user as wizard: A method for early involvement in the design and evaluation of adaptive systems. In *Fifth Workshop on User-Centred Design and Evaluation of Adaptive Systems*, volume 1, pages 460–469, Dublin, Ireland. User-Centred Design and Evaluation of Adaptive Systems.
- [23] Mayhew, D. J. (1991). *Principles and guidelines in software user interface design*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- [24] Mcquiggan, S. W., Mott, B. W., and Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User modeling and user-adapted interaction*, 18(1-2):81–123.
- [25] Moody, J. (2004). Distance education: Why are the attrition rates so high? *Quarterly Review of Distance Education*, 5(3):205.
- [26] Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., and Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education*, 25(2):229–248.
- [27] Okpo, J., Dennis, M., Masthoff, J., Smith, K. A., and Beacham, N. (2016). Exploring requirements for an adaptive exercise selection system. In *PALE 2016: Workshop on Personalization Approaches in Learning Environments*, pages 313–316, Halifax, NS, Canada. CEUR-WS.

- [28] Okpo, J. A., Masthoff, J., Dennis, M., and Beacham, N. (2018). *Adapting exercise selection to performance, effort and self-esteem*. PhD thesis, University of Aberdeen.
- [29] Pratt, D. D., Kelly, M., and Wong, W. S. (1999). Chinese conceptions of 'effective teaching' in hong kong: Towards culturally sensitive evaluation of teaching. *International journal of lifelong education*, 18(4):241–258.
- [30] Rodrigo, M. M. T., Baker, R., Rossi, L., et al. (2013). Student off-task behavior in computer-based learning in the philippines: comparison to prior research in the usa. *Teachers College Record*, 115(10):1–27.
- [31] Smith, K. A., Dennis, M., Masthoff, J., and Tintarev, N. (2019). A methodology for creating and validating psychological stories for conveying and measuring psychological traits. *User Modeling and User-Adapted Interaction*, Early Online.
- [32] Soares, A. M., Farhangmehr, M., and Shoham, A. (2007). Hofstede's dimensions of culture in international marketing studies. *Journal of business research*, 60(3):277–284.
- [33] Taras, V., Roney, J., and Steel, P. (2009). Half a century of measuring culture: Review of approaches, challenges, and limitations based on the analysis of 121 instruments for quantifying culture. *Journal of International Management*, 15(4):357–373.
- [34] Taylor, W. L. (1953). cloze procedure: A new tool for measuring readability. *Journalism Bulletin*, 30(4):415–433.
- [35] Trompenaars, F. and Hampden-Turner, C. (2011). *Riding the waves of culture: Understanding diversity in global business*. Hachette UK, United Kingdom.
- [36] Vasilyeva, E., Pechenizkiy, M., and De Bra, P. (2007). Adaptation of feedback in e-learning system at individual and group level. In *11th International Conference on User Modeling*, volume 1, pages 49–56.
- [37] Woolfolk, A., Winne, P. H., Perry, N. E., and Shapka, J. (2009). *Educational psychology*. Pearson Boston, MA, Boston USA.
- [38] Yoo, B., Donthu, N., and Lenartowicz, T. (2011). Measuring hofstede's five dimensions of cultural values at the individual level: Development and validation of cvscafe. *Journal of International Consumer Marketing*, 23(3-4):193–210.