Ranking of Causes of Failure to Infer Implicature in TOEFL-like Based on Gender

Arifuddin Arifuddin

Department of Language Education, University of Mataram, Mataram Lombok, Indonesia

Abstract—Inferring implicature or pragmatic meaning from TOEFL-like relies on gender. The present study aims at exploring: 1) types of implicature question that male and female test-takers failed to answer and 2) causes of failure and their rankings. This case study involved six students. Data were collected with tests, retrospective report, inventory of causes of failure, interview, recording and analysed with Mixed Methods. Some causes of failure and one type of implicature questions that male and female test-takers' failed to answer are different. Males' ranking is Pronunciation > Sentence Complexity > Cultural Value > Colloquial > Context ≈ Discourse Marker > Mishearing ≈ Speed Rate Delivery ≈ Gender; while females' ranking is Cultural Value > Pronunciation > Sentence Complexity > Context > Discourse Marker > Colloquial. Based on the levels, the ranking is Male Low > Male Medium > Male High > Female Low > Female Medium ≈ Female High.

Index Terms—implicature, TOEFL-like, gender, failure, ranking, discourse marker

I. INTRODUCTION

There is no interactive conversation in the absence of shared knowledge or common ground. Conversation as an example of an interactive verbal communication requires relevance between speaker's intent (or pragmatic meaning) and listener's inference. In a conversation, the speaker generally expresses an utterance which requires the hearer to infer the intended (or pragmatic) 'meaning' or intention commonly called conversational implicature. Arifuddin and Susanto (2012) express that inferring conversational implicatures could be difficult for foreign language learners. Such a difficulty leads to the pragmatic failure. Meanwhile, inferencing ability as an aspect of pragmatic competence is required to raise English language proficiency.

English language proficiency is frequently tested with an instrument called Test of English as a Foreign Language (henceforth TOEFL). Ichsan (1993) tested English language proficiency of his students and found that it is low and listening comprehension is the most difficult skill. Although no detailed information about what makes listening skill low, the study reveals a description of English language proficiency of some English language students in Indonesia.

In relation to teachers' capability, Jalal, et al. (2009) suggest that if the premise that qualified teachers produce qualified students is acceptable, then the poor achievements of students can be attributed to the poor quality of teachers. The achievement of Indonesian students may reflect the lower quality of teachers in Indonesia relative to other countries.

Empirically, a really risky condition indicating low quality of English language learning in English language department in certain higher educations is found. In the 'Visiting Lecture' held at State University of Surabaya on November 20, 2010, Saukah reports that there appears to be a critical problem in English Language Education Program.

The quality of English language teachers in Indonesia is inseparable from their educational background. Saukah (2000) found that English language proficiency of higher education teachers is low, and they do not have good command of English language. A great many English language teachers and even some lecturers show 'insufficient' proficiency. The mean TOEFL scores of English language lecturers in Indonesia is only 390.50, far lower than those of the third year Senior High School students, 435. Among the three sections, Listening section is the most difficult one. This finding is consistent with the one reported by ETS (1997). Based on the data summary, means scores of Listening section, Structure and Written Expression and Reading Comprehension are 63.7, 69.7 and 69.1 respectively. Such a similar condition also appears in Japan. Nishino and Watanabe (2008) found that in Japan, many secondary school English language teachers show low English proficiency on the TOEFL.

According to Saukah (2000), there are a plenty of English language undergraduates who do not show good command of English language. They are not qualified as models in English language teaching. A similar situation occured at other English education in Indonesia. TOEFL-like studies conducted by Arifuddin and Sujana(2004) and Sujana, et al.(2003) showed that mean scores of English proficiency of the 'senior' students of English Education Program at University of Mataram (henceforth EEPUM) is 437, lower intermediate level, far lower than the minimum standard TOEFL-like score of 500 required for graduation. This is 'unique.' Interestingly, their mean score of Listening section is 462, higher than the other two sections. They could only answer correctly 46% of 50 items of listening section involving Short Conversation (Part A) 30 items, Longer Conversation (Part B) 8 items and Monolog/Talk (Part C) 12 items. So, the contribution of Short Conversation (Part A) which assesses pragmatic meanings, implicature in particular, might be high to the low score of both Listening section Parts A and B and TOEFL as a whole. It is probable that one of the causes of low total English proficiency could be low ability in implicature inference from short conversations. Blight

(2002) states that understanding pragmatics or implicatures is still problematic for EFL learners and it hinders proficiency. As a proficiency test, TOEFL generally contains pragmatics (McGormley, http://rcah.msu.edu/language). Unfortunately, TOEFL publishers and ETS never explicitly inform about common barriers of listening comprehension in Paper-based TOEFL.

Regarding content or topics, Part A Listening section of TOEFL-like is gender-free or gender-neutral. However, based on data summaries, for example, ETS test and score data summaries January 2011-December 2011 (ETS, 2012) and January 2008-December 2008 (ETS Researcher, 2008), TOEFL scores are inseparable from test-takers' gender. The mean scores and standard deviations of each section (or skill) according to 'gender' are different. All standard deviations of males' mean scores are higher than those of females. Almost every year ETS reports the means scores of male and female test-takers. It implies that ETS realizes that gender potentially affects test-takers' TOEFL scores. The ETS's concern with the role of gender in English proficiency is a reasonable basis for investigating whether the emergence of types and rankings of causes of failure in inferring implicature from Part A of Listening section of TOEFL-like rely on test-takers' gender.

To guarantee the availability of data and to determine whether implicature comprehension of Part A of Listening Comprehension of TOEFL-like between male and female students of EEPUM is different or not, two preliminary studies were done, on November 11, 2011 attended by 20 males and 32 females and on March 1, 2012 by 30 males and 32 females. These studies aimed at determining whether male students obtain different means correct answers of inferring implicatures by using Part A of Listening section of TOEFL-like. Based on the test administered in the preliminary studies, it indicates that male and female students showed different means of correct answers of implicature inference. Female test-takers' comprehension is better than that of male test-takers. On the basis of their mean scores of pragmatic comprehension, the majority of the test-takers show low ability in inferring implicatures from TOEFL-like. Male test takers achieved far lower mean scores of correct answers than those of females. This finding is definitely relevant to the objectives of the present study.

The difference of mean scores of implicature comprehension of males and females may be related to their preference in language use. It is found that females are prone to produce indirect speech and males direct speech acts in their verbal communication (Mckelvie, 2000; Mulac, Bradac, and Gibbons, 2001). Accordingly, it is *assumed* that: a) female test-takers face fewer causes of failure to understand implicatures from short conversations than do male test-takers and b) Inferring implicatures from short conversations in TOEFL-like needs the test-takers' schema and familiarity with the contexts of indirect speech.

The present study is worth conducting because, as Grujicic-Alatriste (2008) notes, studies focusing on relation between gender and implicatures have not been extensively conducted. Yappy (2002) notes that not much study concerning language functions or pragmatics, particularly used in TOEFL listening section, has been done. That is why, Thijittang and Le (2009) suggest that more research on pragmatics need to be done. A limited number of research findings are reported. For example, it proved that gender influences aural pragmatic understanding (Amin, 2003; Cedar, 2006; Dykstra, 2006) or implicature inference (Barati and Biria, 2011; Cocco and Ervas, 2012). More specifically, women outperformed men in answering inferential comprehension question of aural proficiency tests (John et al., 2003; Farashayian and Hua, 2012). *However*, none of the studies explored the *types* and *rankings* of causes of failure in inferring implicature from short conversations of TOEFL-like on the basis of gender. This is a *gap*. Therefore, it is important to explore them. This is the *novelty* and *authenticity* of the present study.

As mentioned earlier, despite their 'seniority' in terms of semester, the EFL learners' listening comprehension is low. Nissan et al.'s (1996) study shows that TOEFL short dialogues with implicit ideas are difficult for test-takers to infer. In relation to pragmatic failure, the problem is whether male and female test-takers experience different types of causes of failure in each type of implicature question when doing Part A Listening Comprehension of TOEFL-like and how Brown & Yule's(1983) ranking or theory differs from the rankings of causes of failure obtained from the present study. To solve this problem, an investigative study which attempts to reveal gender-specific causes and *rankings* of causes of failure in inferring implicature from Part A Listening section of TOEFL-like has been be conducted. Considering the scope, the obtained rankings are compared with three out of four factors affecting listening comprehension that Brown and Yule (1983) mention. Boyle's (1984) and Rubin's (1994) rankings and causes of failure are used as reference for categorization.

II. THE RESEARCH AIM AND THE QUESTIONS

The present study aims at exploring the:
1. gender-specific causes of failure
2. rankings of causes of failure in inferring implicatures in TOEFL-like based on gender.
The present study pursued the following research questions:
1. What types of implicature question do male test-takers fail to answer? Why?
2. What types of implicature question do female test-takers fail to answer? Why?
3. What are the rankings of causes of failure to infer implicature based on gender?

III. THE SIGNIFICANCE OF THE RESEARCH

Theoretically, the findings clarify Brown and Yule's (1983) 'ranking' and pedagogically, they are meaningful reference for the remedy of EFL learners' aural pragmatic failures.

Pedagogically, the types of implicature question failed to be answered, gender-specific and new rankings of causes of failure are meaningful reference for the remedy of EFL learners' aural pragmatic failures. The application of this approach requires: a) Dedicated teachers with sufficient pragmatic competency responsible for diagnosing students' difficulties and doing remedy, b) English teachers and curriculum developers who are competent in both TOEFL strategies and pragmatics because the nucleus of Communicative Approach or Communicative Language Teaching is *language use* or pragmatics, c) The enthusiasm of the English language educations to introduce TOEFL-like to their learners and d) English language teachers to consider learners' characteristics or, for instance, gender or sex differences. It is recommended that teachers consider EFL learners' differences and treat them as individuals or, at least, as groups, not as a collection of homogeneous creatures.

Since the subjects of the present study are not so familiar with the interlanguage pragmatic content of the authentic proficiency test, TOEFL-like in particular, it is suggested that test developers modify or adapt the pragmatic orientation of the test to suit Indonesian English language learners' schema.

IV. LITERATURE REVIEW

In daily communication, it is assumed that females tend to use more indirect utterances or implicit ideas than males do. This assumption is supported by Markovic's (2007) statement that no one would argue than men use more direct speech constructions than do women because of innate differences between the sexes. It is also relevant to Lindblom's (2010:105) statement indicating that women's speech is more indirect than men's. Similarly, Yule (1996) notes that some researchers have pointed to a preference among women for indirect speech acts rather than the direct speech act.

Empirically, a recent study conducted by Karjo (2011) shows that there are gender differences in production and comprehension of pragmatic meaning. Females are more pragmatic than males. This finding is relevant to Ning et al.'s (2010) research finding indicating that women's speech tendsto reflect pragmatic content than men's.

The theories used are Miller's (2001) and Goldkuhl's (2004) *Typology of Implicature Questions* was used to identify the types of implicature questions asked by the narrators in short conversations in Part A Listening section of TOEFL-like. Brown and Yule's (1983) and Rubin's (1994) theories regarding the *factors which affect listening comprehension* were used to identify and categorize the causes of failure as the basis for their ranking. Meanwhile, Cohen (2012), Cocco and Ervas (2012), John et al. (2003), Amin (2003), Barati and Biria (2011) &Farashayian and Hua (2012) regarding gender and implicature (or pragmatic) comprehension are among the research findings used as references. The research findings and theories are inseparable from the rationale and the design of the present study. Their relationship is illustrated in the following schematic *theoretical framework*.

The fact is that English language proficiency of *senior* students of EEPUM measured with TOEFL-like is low. Preliminary studies indicate that one of the causes is low ability in inferring implicatures from TOEFL-like. Based on some theories and studies, gender affects implicature (or pragmatic) understanding. Theoretically, the present study refers to those theories and research findings. Unfortunately, factors or causes failure in inferring implicature from TOEFL-like and their rankings on the basis of gender have not been extensively explored. This is a *gap*. Therefore, a case study using Mixed methods was employed to investigate the causes of failure for each type of implicature question and gender-specific causes. The data are used as the basis for the formulation of rankings of causes of failure in inferring implicatures from TOEFL-like. The types of cause for each implicature question, gender-specific causes & rankings of causes of failure rely on gender. Theoretically, the findings clarify Brown and Yule's (1983) 'ranking' and pedagogically, they are meaningful reference for the remedy of EFL learners' aural pragmatic failures. Next is the theoretical framework of the present study.
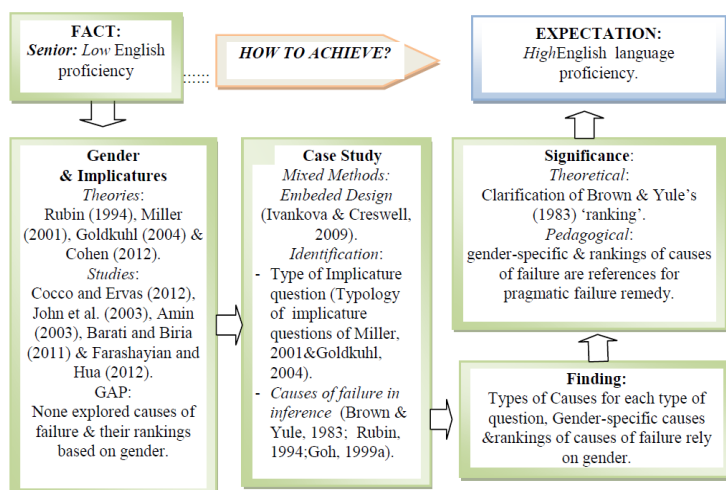
V. METHODOLOGY

A. Subjects of the Research

This case study involved six students of English Education Program University of Mataram selected based on the result of the test in the preliminary study. Implicature questions asked in Part A Listening comprehension TOEFL-like are the materials of present study.

B. Instruments for Data Collection

Data about causes of failure to infer implicature were collected with tests (Part A TOEFL-like), retrospective report, inventory of causes of failure, interview, recording and human instrument.

C. Procedures

The present study was conducted according to the following phases: a) Preliminary studies; b) Subject selection using Part A Listening TOEFL-like; c) Data Collection; d) Data Analysis; and e) Verification of the findings. Note that the identification and categorisation of causes of failure and causes of failure employed Miller's (2001) and Goldkuhl's (2004) Typology of Implicature and Brown and Yule's (1983) and Rubin's (1994).

D. Data Analysis

Data were analysed with Mixed Methods: Embeded Design. Qualitative analysis was done using Yin's (2011) iterative qualitative analysis integrated with Descriptive Quantitative: Frequency.

VI. FINDINGS

The description of causes of failure for each type of implicature question related to gender-specific causes presented in Table 1belowaims at providing the answer for Research Question 1, "How do the rankings of causes of failure of male test-takers differ from those of female test-takers?"

TABLE 1.
GENDER-SPECIFIC CAUSES OF FAILURE

	Gender-specific Causes	
	Male	Female
Types of Question: Causes of Failure	*Meaning/Intent question*: Mishearing, Speed Rate/Delivery& Context	*Meaning/Intent question*: Discourse Marker
	Inference question: Colloquial	*Inference question*: Not identified
	Reference question: Context	*Reference question*: Sentence Complexity
	Deixis question: Not identified	*Deixis question*: Sentence Complexity, Cultural Value & Context
	Presupposition question: Pronunciation, Gender & Discourse Marker	*Presupposition question*: Context
	Predictable action question: Not identified	*Predictable action question*: Context
Overall	Mishearing,Speed Rate/Delivery& Gender	Not identified

Based on TABLE 1, gender or sex difference is obviously apparent because some gender-specific causes of failure are identified. It is described below.

In *Meaning/Intent* question, gender-specific causes of failure are found. Mishearing, Speed Rate/Delivery and Context appear as *gender-specific: male,* while Discourse Marker appears as *Gender-specific: female.* In Inference question, Colloquial is identified as *gender-specific*: *male*. In Reference question, Context appears as *gender-specific*: *male* and Sentence Complexity as *gender-specific*: *female*. In Deixis question, Sentence Complexity, Cultural Value and Context are identified as *gender-specific: female.* Gender-specific is identified in Presupposition question. 'Man's voice is easier.' Gender, Pronunciation and Discourse Marker appear as *gender-specific: male.* For female test-takers, Context is identified as *gender-specific*: *female*. *Predictable action question* is answered incorrectly only by female test-takers, so that this type of implicature question is *gender-specific*: *female.*

Meaning question is the most dominant cause of failure. Of all types of question, except for 'Deixis' and 'Predict,' male test-takers faced broader coverage of cause of failure than female test-takers. The difference of total frequency of causes of failure in Meaning/Intent question between male test-takers and female test-takers is extremely high, 28:11. It means that males are superior to females in terms of level of causes of failure. This is relevant to some theories claiming that female test-takers outperform male test-takers in listening comprehension.

Male test-takers faced broader causes of failure in Meaning, Inference and Presupposition questions than female test-takers did. However, they faced narrower coverage of failure in Deixis and Predicatable action questions than female test-takers did. Nevertheless, male test-takers and female test-takers shared the same coverage of causes of failure in Meaning and Reference questions. Thus, based on the total causes of failure in all types of implicature question, *gender-*

specific: male is superior to *gender-specific*: *female*. The rankings based on the levels (or frequencies) of causes of failure related to gender-specific causes are briefly described below.

For male test-takers, the highest level cause of failure is Pronunciation (*60*) and the lowest one is Gender (*1*). Male test-takers cover far broader causes of failure than female test-takers. Only did male test-takers face Mishearing (1), Speech Rate/Delivery (*1*) and Gender (*1*) as causes of failure. Besides Pronunciation, Sentence Complexity (*14*) and Cultural Value (*9*) also place high level of causes. Colloquial (*4*), Context (*3*) and Discourse Marker (*3*) are in the middle position. Context and Discourse Marker share the same level of causes, so do Mishearing, Speech Rate/Delivery and Gender. Mishearing, Speed Rate/Delivery and Gender are *gender-specific*: *male*. Below is male test-takers' ranking:

$Pr> SC > CV > COL > Cont \approx MarkRh > MisH \approx SRD \approx Gen.$

For female test-takers, Cultural Value (*20*) is the in the highest level of cause of failure, while Colloquial (*1*)is in the lowest one. Pronunciation (*18*) and Sentence Complexity (*18*) are also in the high position. Like male test-takers, female test-takers also place Context (*11*) in the middle position. However, as a whole, no *gender-specific: female* appears. It means that female test-takers experienced less diverse causes of failure than male test-takers. The following is the female test-takers' ranking:

$CV > Pr > SC > Cont > MarkRh > COL.$

Regarding superiority, male test-takers are superior to female test-takers in Pronunciation (*60*), Colloquial (*4*), Mishearing (*1*), Speech Rate/Delivery (*1*), Discourse Marker (*3*) and Gender (*1*). In Sentence Complexity (*15*), Cultural Value (*20*) and Context (*11*), female test-takers achieved higher level of causes of failure than male test-takers did. Accordingly, the rankings are as follows:

Male>Female: Pr, COL, MisH, SRD, MarkRh, Gen; while *Male<Female: SC, CV, Cont.*

In addition, related to the levels of the subjects, high (H), medium (M) and low (L); the order of the three levels of subjects based on frequencies of causes is Male Low (*55*), Male Medium (*26*), Male High (*19*), Female Low (*16*), Female Medium (*15*) and Female High (*15*). Interestingly, the three levels of female subjects show relatively comparable number of frequencies. Accordingly, the ranking of causes of failure based on their frequencies is as follows:

Male L > Male M > Male H > Female L > Female M ≈ Female H.

It indicates that all three levels of male subjects experienced more causes of failure than those of three levels of female subjects.

To sum up, in the context of the present study, male test-takers and female test-takers differ in terms of: 1) emergence of causes to a peculiar sex (or natural gender) (gender-specific) and 2) coverages and levels of causes of failure. *Male test-takers experience broader coverage of gender-specific causes of failure in implicature inference of short conversations than female test-takers do.* In other words, female test-takers experienced less diverse causes of failure than male test-takers. This is a new finding (or *theory*) generated from the present study.

The presentation of coverage and order of causes of failure of male test-takers and female test-takers in tabular form is to show how the obtained rankings are derived and how they differ from Brown and Yule's (1983) 'ranking'. The causes of failure as sub-factors are categorised under each factor as appropriate. This presentation is aimed at answering.

Research Question 2, "How do the rankings of categorised causes of failure in inferring implicatures obtained in the present study differ from Brown and Yule's (1983) 'ranking'?

There are three components which are compared with Brown and Yule's (1983) 'ranking':

1) Total (Sub-factors + Frequency);

2) Sub-factors

3) Frequencies.

This comparison reveals new ranking(s) of factors affecting the failure in communicative listening, implicature inference of Part A Listening Comprehension of TOEFL-like in particular.

The coverage and sequence factors which afffect failure in implicature inference are based on Brown and Yule (1983), Underwood (1989), Rubin (1994), Van Duzer (1997), Goh (1999a), Van Duzer (2003), Goh and Taib (2006), Graham (2006), Kijpoonphol (2010), and But *et al.* (2010). Some of the specific factors (or features) mentioned by the scholars are overlapped. In categorising the factors or causes of failure, the researcher carefully examined the characteristic similarity of the specific causes.

The detailed coverage and order of causes of failure that male test-takers and female-test-takers experienced used as the bases for ranking formulation are written. Furthermore, the total frequency of each cause of failure is placed under the column of sub-factors of appropriate factor (*see* TABLE 2). In Table 2, the categorisation of causes or factors which affect implicature inference follows Brown and Yule's (1983) 'ranking'. Besides, the taxonomy of subfactors or features under each factor is based on categorised factors affecting listening of Rubin (1994), Van Duzer (1997), Goh (1999a), Goh and Taib (2006) and Kijpoonphol (2010).

TABLE 2.
COVERAGE AND ORDER OF MALE TEST-TAKERS' AND FEMALE TEST-TAKERS' CAUSES OF FAILURE

Male Test-takers		**Female Test-takers**	
Factor	Sub-factor	Factor	Sub-factor
Listener	CV = 9, Cont= 3 and MisH = 1	Listener	CV =20 and Cont= 11
	Sub-factor: 3 Frequency: 13		Sub-factor : 2 Frequency: 31
Speaker	Pr (Voice, BS, Acc) = 60 COL = 4, Sex/Gen= 1 & SRD = 1.	Speaker	Pr (Voice, BS, Acc) = 18 &COL = 1.
	Sub-factor: 6 Frequency: 66		Sub-factor: 4 Frequency: 19
Content	SC = 14 & MarkRh = 3	Content	SC = 16 &MarkRh= 2.
	Sub-factor: 2 Frequency: 17		Sub-factor: 2 Frequency: 18
Visual Support	*Not involved*.	Visual Support	*Not involved*.
Total	Sub-factors (Coverage): 11 Frequency(Order): 96	Total	Sub-factors (Coverage): 8 Frequency (Order) : 68
Male Test-takers' Ranking: Coverage: *Speaker > Listener > Content* Order: *Speaker > Content > Listener* Coverage + Order: *Speaker > Content > Listener*		Female Test-takers' Ranking: Coverage: *Speaker > Listener ≈ Content* Order: *Listener > Speaker > Content* Coverage + Order: *Listener > Speaker > Content*	
Brown & Yule's (1983) (non-gender-based): *Listener > Speaker > Content*			

Codes:

CV= *Cultural Value*, Cont= *Context*, MisH= *Mishearing*, Pr= *Pronunciation*, BS= *Blended sounds*, Acc= *Accent*, COL= Colloquial, Gen= Gender, SRD= *Speech Rate/Delivery*, SC= *Sentence Complexity*, MarkRh= *Discourse Marker*.

VII. DISCUSSION

Based on the summary data presented in Table 2, how male test-takers' rankings of causes of failure differ from female test-takers' is summarised below:

a) As a whole, the number of sub-factors under each factor and the total of sub-factors and frequency of causes of male test-takers' failure are higher than those of female test-takers'. It means that male test-takers experienced more causes of failure in implicature inference than female test-takers did.

b) In terms of 'coverage', the primary cause of failure is 'Speaker'. However, female test-takers were so sure that Content and their own listening competency (as Listener) share the same degree of causes of failure, while male test-takers admitted that Content is not a significant cause of failure. It means that male test-takers comprehend Content better than female test-takers do.

c) Of the three female test-takers' rankings, two rankings place Listener as a cause of failure in listening to short conversations, while male test-takers consider that the main factor which contributes to their failure is Speaker, not they themselves as Listener. The fact that male test-takers' and female test-takers' rankings are different proves that there is an effect of gender on the failure in inferring implicatures from Part A Listening Comprehension of TOEFL-like.

d) Brown & Yule's (1983) 'ranking', *Listener > Speaker > Content,* is different from all male test-takers' rankings (Coverage, Order and Coverage + Order) and one of three female test-takers' rankings (Coverage). Only female test-takers' Order and Coverage + Order rankings are the same as Brown and Yule's (1983) 'ranking'.

The emergence of gender-specific causes of failure in each type of implicature question is described as follows. In *Meaning/Intent* question, gender-specific causes of failure is identified. Mishearing and Speed Rate/Delivery appear as *gender-specific: male,* while Discourse Marker and Context are found to be *gender-specific: female.* In Inference question, Colloquial is identified as *gender-specific: male*. This finding is relevant to Eisenstein's (1982) research finding which shows that females consistently and significantly outperformed males in discriminating among different American English sounds.

For female test-takers, Sentence Complexity, Cultural Value and Context are identified as *gender-specific: female* in Deixis question. Gender-specific is also identified in Presupposition question. 'Man's voice is easier.' Gender, Pronunciation and Discourse Marker appear as *gender-specific: male.* For female test-takers, Context is identified as *gender-specific: female.* Predictable action question is answered incorrectly only by female test-takers, so that this type of implicature question is *gender-specific: female.* In conclusion, males dominate female in both *gender-specific.* This finding is relevant to Zaidi's (2010) note reporting that some neurological studies have shown that men use the left side of their brains when they listen to someone speaking, whereas women use both sides. Similarly, Lenarz, et al. (2012) also state that women show greater bilateral activity in language. This means that women may face fewer causes of failure in listening than do men.

Meaning question is the most dominant cause of failure. Of all types of question, except for 'Deixis' and 'Predict,' male test-takers faced broader coverage of cause of failure than female test-takers. The difference of total frequency of

causes of failure in Meaning/Intent question between male test-takers and female test-takers is extremely high, 28:11. This finding is relevant to Zaidi's (2010) note reporting that some studies have shown that men listen better than men.

Male test-takers faced broader causes of failure in Meaning, Inference and Presupposition questions than female test-takers did. Male test-takers faced narrower coverage of failure in Deixis and Predicatable action questions than female test-takers did. Male test-takers and female test-takers shared the same coverage of causes of failure in Meaning and Reference questions.

Based on their levels, the rankings of causes of failure of male test-takers and female test-takers are described as follows.In male test-takers' ranking,

$Pr > SC > CV > COL > Cont \approx MarkRh > MisH \approx SRD \approx Gen,$

the highest level cause of failure is Pronunciation and the lowest ones are Mishearing, Speech Rate/Delivery and Gender. This is in line with Larsen-Freeman and Long (1991) hypothesis that females may possess better listening skills than do males.

To female test-takers, Cultural Value is the in the highest level of cause of failure, while Colloquial is in the lowest one. Pronunciation and Sentence Complexity are also in the high position. Like male test-takers, female test-takers place Context in the middle position. However, as a whole, no *gender-specific*: *female* appears. It means that female experience less diverse causes of failure than male test-takers. In female test-takers' ranking,

$CV > Pr > SC > Cont > MarkRh > COL,$

Cultural Value is the in the highest level of cause of failure, while Colloquial is in the lowest one. Pronunciation and Sentence Complexity are also in the high position. Like male test-takers, female test-takers also place Context in the middle position. However, as a whole, no *gender-specific: female* appears. It means that female test-takers experienced less diverse causes of failure than male test-takers.

In addition, regarding the levels of the subjects (High, Medium and Low), the holistic ranking based on the frequencies of causes of failure of is as follows:

Male L > Male M > Male H > Female L > Female M ≈ Female H.

In relation to superiority, male test-takers are superior to female test-takers in Pronunciation, Colloquial, Mishearing, Speech Rate/Delivery, Discourse Marker and Gender. In Sentence Complexity, Cultural Value and Context, female test-takers achieved higher level of causes of failure than male test-takers did. The rankings are as follows:

Male>Female: Pr, COL, MisH, SRD, MarkRh, Gen; while

Male<Female: SC, CV, Cont.

In conclusion, in the context of the present study, male test-takers and female test-takers differ in terms of: 1) emergence of causes to a peculiar sex (or natural gender) (gender-specific) and 2) coverages and levels of causes of failure. *Male test-takers experience broader coverage of gender-specific causes of failure in implicature inference of short conversations than female test-takers do.* Female test-takers experienced less diverse causes of failure than male test-takers. This finding is consistent with Barati and Biria's (2011) and Cocco and Ervas's (2012) studies indicating that gender influences implicature inference or pragmatic understanding.

In relation to Research Question 2*,* there are three components which are compared with Brown and Yule's (1983) 'ranking': 1) Total (Sub-factors + Frequency); 2) Sub-factors and 3) Frequencies. This comparison reveals new ranking(s) of factors affecting the failure in communicative listening, implicature inference of Part A Listening Comprehension of TOEFL-like, in particular.

As a whole, the number of sub-factors under each factor and the total of sub-factors and frequencies of causes of male test-takers' failure are higher than those of female test-takers'. It means that male test-takers experienced more causes of failure in implicature inference than female test-takers did. This finding is relevant to Farashayian and Hua's (2012) and John et al.'s (2003) studies which show that women outperformed men in answering inferential comprehension questions of auditory discourses in proficiency tests.

Regarding the 'coverage', both male test-takers and female test-takers experienced 'Speaker' as the primary cause of failure. However, female test-takers were so sure that Content and their own listening competency (as Listener) share the same degree of causes of failure, while male test-takers admitted that Content is not a significant cause of failure. It means that male test-takers comprehend Content better than female test-takers do.

Of the three female test-takers' rankings, two rankings place Listener as a cause of failure in listening to short conversations, while male test-takers consider that the main factor which contributes to their failure is Speaker, not they themselves as Listener.

The fact that male test-takers' and female test-takers' rankings are different proves that there is an effect of gender on the failure in inferring implicatures from Part A Listening Comprehension of TOEFL-like. Consequently, the research findings which show that listening comprehension(Amin, 2003) and implicatureor pragmatic understanding (John et al., 2003; Barati and Biria, 2011; Cocco and Ervas, 2012; and Farashayian and Hua, 2012) between men and women are different is *confirmed*. Thus, gender is one of the factors which affect the success or failure in inferring implicature from short conversations.

The following are three previous studies and one proposed theory which all report *coverages* of factors which affect communicative listening comprehension. Uniquely, the *coverages* and *orders* of factors of the three studies are different. The emergence of the diverse *rankings* provides scientific opportunity for conducting further comparative investigations.

The present study provides the detail of gender-specific causes and rankings of causes of failure that learners experienced when inferring implicatures from short dialogues. Theoretically, this finding: a) is useful for the clarification of categorisations formulated in the previous studies, b) reveals gender-specific and ranks of causes of failure in implicature inference formulated based on gender; and c) is the basis for clarifying Brown and Yule's (1983) 'ranking' of factors affecting listening.

The finding of the present study meaningfully clarifies Brown and Yule's (1983) 'ranking' of factors which affect listening comprehension. Indirectly, the new finding yields further information for Amin's (2003) study, Goh's (1999a) and Chang's (2009) studies. Hopefully, in reference to these new rankings, EFL learners will not 'overgeneralize' those findings, Brown and Yule's (1983) 'ranking of factors', in particular.

Pedagogically, the types of implicature question failed to be answered, gender-specific and new rankings of causes of failure are meaningful reference for the remedy of EFL learners' aural pragmatic failures. The application of this approach requires: a) Dedicated teachers with sufficient pragmatic competency responsible for diagnosing students' difficulties and doing remedy, b) English teachers and curriculum developers who are competent in both TOEFL strategies and pragmatics because the nucleus of Communicative Approach or Communicative Language Teaching is *language use* or pragmatics, c) The enthusiasm of the English language educations to introduce TOEFL-like to their learners and d) English language teachers to accommodate learners' characteristics or, for instance, gender or sex differences. It is recommended that teachers consider EFL learners' differences and treat them as individuals or, at least, as groups, not as a collection of homogeneous creatures.

Since the subjects of the present study are not so familiar with the interlanguage pragmatic content of the authentic proficiency test, TOEFL-like in particular, it is suggested that test developers carefully create short conversations which contain more universal pragmatics.

The findings of the present study will be useful for language centres' TOEFL instructors in selecting and designing TOEFL listening materials, and applying appropriate teaching and learning strategies which facilitate comparable trainees to effectively comprehend pragmatic capacity. Besides, it is a good idea for a language centre to recognise the trainees' difficulties in raising their TOEFL scores. Basically, TOEFL score should be raised through self-study or independent learning. The trainees need to apply independent learning which focuses on their listening problems or weaknesses. The causes of failure in listening obtained from the present study are good references. By so doing, it is expected that the trainees' English proficiency be improved. Language Centre University of Mataram, for instance, is responsible for designing TOEFL preparation materials for those who apply for further study. Familiarity with the types of implicature questions could help the trainees or test-takers comprehend pragmatic meanings. To achieve high score in Part A Listening Comprehension of TOEFL-like, it is essential that the test-takers be familiar with the types of questions and specification of content of the conversations.

VIII. CONCLUSION

It is concluded that: 1) male test-takers dominate gender-specific causes of failure. To male test-takers, the highest level cause of failure is Pronunciation and the lowest one is Gender. Pronunciation, Sentence Complexity and Cultural Value also place high level of causes. Context is in the middle position. Context and Discourse Marker share the same level of causes, so do Mishearing, Speech Rate/Delivery and Gender. Mishearing, Speed Rate/Delivery and Gender are *gender-specific*: *male*. Accordingly, male test-takers' ranking is:

Pr> SC > CV > COL >Cont ≈ MarkRh>MisH ≈ SRD ≈ Gen.

Female test-takers' ranking is:

CV >Pr> SC >Cont>MarkRh> COL.

Male test-takers are superior to female test-takers in Pronunciation, Colloquial, Mishearing, Speech rate/Delivery, Discourse Marker and Gender. In Sentence Complexity, Cultural Value and Context, female test-takers achieved higher level of causes of failure than male test-takers did. The rankings are as follows:

Male>Female: Pr, COL, MisH, SRD, MarkRh, Gen; while *Male<Female: SC, CV, Cont.*

These rankings indicate that male test-takers experience broader coverage of causes of failure in implicature inference of short conversations than female test-takers do. This is a *new finding* from the present study.

In addition, the holistic ranking of causes based on the frequencies of causes of failure of the three levels of subjects -- High, Medium and Low -- is as follows:

Male L > Male M > Male H > Female L > Female M ≈ Female H.

Thus, all levels of male subjects show higher ranks of causes of failure in implicature inference than those of female subjects. This is another *finding*. As a whole, the number of sub-factors under each factor and the total of sub-factors and frequencies of causes of male test-takers' failure are higher than those of female test-takers'. It means that male test-takers experienced more causes of failure in implicature inference than female test-takers did. In terms of *coverage,* both male test-takers and female test-takers experienced 'Speaker' as the dominant cause of failure.

The rankings formulated based on the results of the present study differ from Brown and Yule's (1983) 'ranking'. The order of the three factors in their ranking formulated as *Listener > Speaker > Content* is different from all male test-takers' rankings: a) Coverage: *Speaker > Listener > Content*, b) Order: *Speaker > Content > Listener* and c) Coverage + Order: *Speaker > Content > Listener*) and one of the three female test-takers' rankings, that is, Coverage:

Speaker > Listener ≈ Content. Only female test-takers' Order and Coverage + Order rankings are the same as Brown and Yule's (1983) 'ranking'. Consequently, limited to causes of failure in inferring implicatures from short conversations in TOEFL-like, their 'ranking' or 'theory' is now *clarified*. This finding proves that gender is one of the factors which affect failure in inferring implicature from short conversations in TOEFL-like.

It is urgent to do further studies which explore detailed: a) example(s) of each specific cause of failure in inferring implicature, e.g. example(s) of intercultural values and b) rankings of causes of failure for each type of implicature question of each level of subjects.

REFERENCES

[1] Amin, K. (2003). The effect of speaker's and listener's gender on L2 listening comprehension. *Indian Journal of Applied Linguistics* 29.1, 99-107.

[2] Arifuddin and I.M. Sujana. (2004). Kecakapan berbahasan Inggris mahasiswa senior Prorgram Studi BahasaInggris LPTK di NTB. Jurnal Penelitian Universitas Mataram, Februari 2004.

[3] Arifuddin and Susanto. (2012). Gender-based failure to infer implicatures from TOEFL-like listening."*International Journal of Learning and Development* 2.6, 62-72. http://dx.doi.org/10.5296/ijld.v2i6.2681.

[4] Barati, L. and R. Biria. (2011). The impact of first language intonational clue selection on second language comprehension. *Open Journal of Modern Linguistics* 1.2, 33-38. http://dx.doi.org/10.4236/ojml.2011.12005.

[5] Boyle, J.P. (1984). Factors affecting listening comprehension. *ELT Journal* 38.1, 34-38. http://dx.doi.org/10.1093/elt/38.1.34.

[6] Brown, G. and G. Yule. (1983). Teaching the spoken language. Cambridge: Cambridge University Press.

[7] Butt, M.M, Sharif, M.M., Naseer-ud-Din, M., Hussain, I., Khan, F., and Ayesha, U. 2010."Listening comprehension problems among the students: A case study of three govt. boy's higher secondary schools."*European Journal of Social Sciences*18(2): 311-315.

[8] Cedar, P. (2006). Thai and American responses to compliments in English."*The Linguistics Journal* 1.2, 6-28.

[9] Cocco, R. and F. Ervas. (2012). Gender stereotypes and figurative language comprehension. *Hummana Mente Journal of Philosophical Studies* 22, 43-56.

[10] Dykstra, L.K. (2006). On pragmatic perception: Do learners of Russian perceive the socio-cultural weight of the address pronouns? Unpublished Ph.D. Dissertation University of Iowa, USA.

[11] Eisenstein, M. (1982). A study of social variation in adult second language acquisition. *Language Learning* 32, 367-92. http://dx.doi.org/10.1111/j.1467-1770.1982.tb00977.x.

[12] ETS Researchers. (2008). Validity evidence supporting the interpretation and use of TOEFL iBT scores. New Jersey: Princeton.

[13] ETS. (1997). TOEFL: Test and score manual. New Jersey: Princeton.

[14] ETS. (2012). 2011-2012 Information and registration bulletin for paper-based testing (PBT). Princeton: New Jersey.

[15] Farashayian, A. and T.K. Hu. (2012). On the relationship between pragmatic knowledge and language proficiency among Iranian male and female undergraduate EFL learners. *The Southeast Asian Journal of English Language Studies* 18.1, 33-46.

[16] Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics* 4.1, 17-42.

[17] Goh, C. and Y. Taib. (2006). Metacognitive instruction in listening for young learners. *ELT Journal* 60, 222-232. http://dx.doi.org/10.1093/elt/ccl002.

[18] Ichsan, D. (1993). Tingkat kecakapan berbahasa Inggris mahasiswa program studi Pendidikan Bahasa Inggris Fakultas Keguruan dan Ilmu Pendidikan Universitas Sriwijaya Angkatan tahun 1987 dan tahun 1988. Prosiding Seminar Hasil Penelitian Bidang Kependidikan BKS-PTN Bagian Barat ke-2 (Eds, Utomo, M., Sudrajat, Sujarwo, and Sudarmanto, R.G.)., Bandar Lampung, Lembaga Penelitian Universitas Lampung, hal. 113-125.

[19] Jalal, F., Samani, M., Chang, M.C., Stevenson, R., Ragazt, A.B., and S.D. Negara. (2009). Teacher certification in Indonesia: A strategy for teacher quality improvement. Jakarta: Depdiknas RI.

[20] John, S.F, Lui, M., and R. Tannock. (2003). Children's story retelling and comprehension using new narrative resource. *Canadian Journal of School Psychology* 18.1, 291-113.

[21] Kijpoonphol, W. (2010). An investigation of difficulties related to listening comprehension: A case study of university students. *NIDA Language and Communication Journal*, 30-51.

[22] Lenarz, M., Sonmez, H., Joseph, G., Buchner, A. and T. Lenarz. (2012). Effect of gender on the hearing performance of adult cochlear implant patients. *The Laringoscope* 122, 1126-1129. http://dx.doi.org/10.1002/lary.23214.

[23] McGormley, K. RCAH Michigan State University. Retrieved on 7 July 2011 from http://rach.msu.edu/language.

[24] Mckelvie, B. (2000). The effect of hyper feminity on communication pattern in dating couples. *Dissertation Abstract International* 60, 63-75.

[25] Mulac, S.,Bradac, J., and V.C. Gibbons. (2001). Empirical support for the gender-as-a-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research* 27, 121-152. http://dx.doi.org/10.1093/hcr/27.1.121.

[26] Nishino, T. and M. Watanabe. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly* 42.1, 133-138.

[27] Nissan, S., DeVincenzi, F., and K.L. Tang. (1996). An analysis of factors affecting the difficulty of dialogue items in TOEFL® listening comprehension. TOEFL Research Report 51. New Jersey: Princeton.

[28] Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal* 78.2, 199-221. http://dx.doi.org/10.1111/j.1540-4781.1994.tb02034.x.

[29] Saukah, A. (2000). The English proficiency of the academics of the teacher training and education institutions. *Jurnal Ilmu Pendidikan* 7.1, 67-76.

[30] Sujana, I.M.,Syahrial, E. and E. Fitriana. (2003). Profisiensi bahasa Inggris mahasiswaS1 bahasaInggris FKIP Universitas Mataram dalam Test of English as A Foreign Language (TOEFL). *Jurnal Penelitian Universitas Mataram* 2.3, 14-26.

[31] Thijittang, S. and T. Le. (2009). Gender differences and apologies in English of Thai learners: Pragmatic and sociolinguistic perspectives. *Proceedings of the 5th Biennieal International Gender and Language Association Conference IGALA 5* (Eds. de Bres, J., Holmes, J., Marra, M.), held at Victoria University of Wellington, July 2008, pp. 51-66.

[32] Underwood, M. (1989). Teaching listening. New York: Longman.

[33] Van Duzer, C. (1997). Improving ESL learner's listening skills: At the work place and beyond. ERIC publication February 1997, New York, CAL.

[34] Yappy, S.N. 2002."An analysis of language functions in short talks of TOEFL listening section: Towards pedagogical and theoretical implications."*Magister Scientiae Journal* No. 11 March 2002.

[35] Zaidi, J.F. (2010). Gender differences in human brain: An overview. *The Open Anatomy* 2, 37-55. http://dx.doi.org/10.2174/1877609401002010037.

Arifuddin Arifuddin is an experienced educator. He holds his Ph.D. in Language and Literature Education from the State University of Surabaya Indonesia. His research interest includes TEFL, pragmatics, and language assessment.