**MANGAL RESEARCH JOURNAL**
**Volume** 5, December 2024, **pp:** 1-18
**ISSN:** 2773-7802
**DOI:** https://doi.org/10.3126/mrj.v5i01.73470

## Editorial

# A Systematic Review of Error Gravity Articles: Themes of Leniency, Perceptions, and Hierarchies

**Prof. Laxmi Bahadur Maharjan, PhD**

Editor in Chief
Mangal Research Journal
Email: editors.mrj@gmail.com
laxmi.maharjan@tucded.edu.np

**Abstract**

*This systematic review synthesizes findings from various studies published before 2000 to understand the frameworks and criteria researchers use to assess error gravity. The study aims to explore how different types of errors are evaluated regarding their severity and impact, as well as perceptions in spoken and written language, ultimately establishing hierarchy of errors. I selected thirty-one articles on error gravity in language assessment published from 1969 to 1999 using a purposive sampling method, focusing exclusively on studies that directly address the evaluation of error gravity. The sources for the articles in my study included various journals, dissertations, monographs, ERIC documents, and conference proceedings. I conducted an in-depth study of these articles, resulting in four themes: Perceptual Dichotomy: Leniency vs. Seriousness of Errors, Perception of Errors in Spoken Communication, Perception of Errors in Written Compositions and* Establishing Hierarchies of Error. *The review study revealed that tolerance toward ESL errors has become a distinctive characteristic of native speakers, reflecting perceptual trends closely tied to their native-speaker identity. The study concluded that native speakers perceive errors in a hierarchy based on severity. However, the concept of a "universal hierarchy of errors" lacks conclusive evidence. McCretton and Rider (1993) emphasized that these hierarchies are shaped by evaluators' educational backgrounds, challenging the idea of inherent universality.*

***Keywords:*** Error gravity, hierarchy of errors, leniency, perceptions, perceptual trends

## Introduction

Error perception, leniency, seriousness, and the hierarchy of errors are recurring terminologies in the analytical studies on error gravity. The terms error perception, error evaluation, and error judgment are often used interchangeably by researchers to describe similar concepts. All researchers arrive at a unanimous conclusion that ESL errors are evaluated by native English speaking teachers and non-teachers both leniently in opposition to non-native teachers who evaluated harshly. James (1998, p. 205) explains the purpose of error evaluation in the following way: “to prevent obsession with trivial

errors and give priority to the ones that really matter." However, it has been accepted that there are consistent differences between the way nationals (non-native speakers) and native-speaker teachers mark written exercises (James, 1977). Judging student errors require guidelines for determining their seriousness of gravity. One can speculate that individual teachers of ESL, regardless of their native language, tend to evaluate errors or error types differently (Sheorey, 1986). James seems to support this view when he says that ESL teachers probably 'do refer consistently a criterion of degrees of erroneousness when they mark, even though they do not explicitly formulate these criteria. (p.116). Sheorey (1986) looks interested to find out whether there are any consistent and (statistically) significant differences between the error-gravity perceptions of native and non-native teachers of ESL.

The linguists of different times carried out research on error gravity. Johansson (1978) was one of the earliest to focus on error evaluation and reactions to non-native English. Explaining the relevance of his study, he says that "error evaluation must be considered a largely neglected field of research" (p. 1), although a need to establish error gravity had previously been pointed out by James (1972, p. 76) and Robinson (1973, p. 192), and Bansal (1969) and Olsson (1972, 1973) occasionally attempted to do it in a systematic way. There is a lesser-known term, 'low-gravity errors' (Ilin, 2017), which is used to describe erroneous utterances that are comprehensible and do not cause irritation.

Native speaker teachers, non-teachers, and non-native teacher judges of various ages, genders, qualifications and professions have been involved in evaluating the seriousness of errors at the word, sentence, and composition levels. Previous studies have identified factors such as sex, age, profession, and educational experience as influencing evaluators' judgments (Olsson, 1973; Politzer, 1978; Ensz, 1982). However, as Ludwig (1982) highlights, this area requires further research (Vann, Meyer, & Frederick, p. 433). Efforts have also been made to determine whether correlations exist between error gravity judgments and one or more variables. Notable variables examined include teaching experience (James, 1977; Oliaei & Sahragard, 2013), academic specialization (Roberts & Cimasko, 2008; Santos, 1988; Vann et al., 1984), and respondents' age (Vann et al., 1984) (as cited in Endley, 2016).

Error evaluation is conducted at various levels, including words, sentences, and compositions. Extensive research has been carried out to explore the seriousness of errors made by non-native teachers. Numerous articles have been published examining the perceptions of both native and non-native teachers regarding different types of errors. Error gravity, as a concept, extends beyond written language to include the perception of errors in spoken English. Bansal's (1969) work on the intelligibility of Indian English provides a foundational perspective in this area, focusing on how phonological deviations

impact communication effectiveness. His work parallels later theoretical frameworks on error gravity, making his contribution significant in understanding how phonological errors influence error perception and severity judgments. Many researchers have confined their studies in written errors. There are error evaluation researchers who began to pay greater attention to the linguistic context in which errors appear by abandoning the use of decontextualized sentences as learner language samples. Several earlier studies (Chastain, 1980; Delisle, 1982; Khalil, 1985; Piazza, 1980) used sentences as learner language samples. However, Haswell (1988), Lennon (1991), Santos (1988) and Vann et al. (1984) preferred the use of whole learner written compositions (as cited in Tong, 2000).

Several studies have attempted to establish error hierarchies for specific languages: English (Olsson 1973), German (Politzer 1978), Spanish (Guntermann 1978, Chastain 1980), and French (Piazza 1980). However, at least one group of researchers concluded that a search for a hierarchy of errors is fruitless, stating that "all errors are equally irritating … irritation is directly predictable from the number of errors regardless of the error type of other linguistic aspects" (Albrechtsen, Henriksen, and Faerch, 1980, p. 394, as cited in Vann et al., 1984, p. 428). A hierarchy of error types was more successfully completed by Hultfors (1986). He undertook a profound investigation of error gravity based on the criteria of acceptability and intelligibility, also accounting for the variation of participants and the 'foreigner role' as outlined by Johansson (1978).

McCretton and Rider (1993) suggested there must be a "universal hierarchy of errors". Some researchers noticed that the judges were consistent in the ranking of the error types, which suggests that the categories have some psychological validity (cf. James 1977). One of the explanations for the tendencies in foreign language teachers is that they must have developed a system of principles for the evaluation that guide them in marking students' writing. Vann, Meyer and Fredrick (1984) note that most people see errors "in relative rather than absolute terms" (p. 437), it is therefore sensible to order them according to their importance. All error gravity hierarchies have been adapted for better comparison and organized in the same way – from least severe error types at the bottom to most severe ones at the top. To summarize the results, Rifkin and Roberts (1995) made a review of 28 error gravity studies (1977-1995) and were somewhat disappointed at the inconsistent findings of the error investigations that "make it difficult to point confidently in any one direction and proclaim it the route for improving native/nonnative interaction" (p. 512).

The study seeks to examine the evaluation of various types of errors based on their severity and impact, as well as perceptions of errors in both spoken and written language, with the goal of establishing a hierarchy of errors.

## Methods

This qualitative review study investigates error perception in error gravity research, focusing on themes of leniency, medium (spoken or written) in the evaluation of errors, and hierarchical evaluations. A purposive sampling method was employed to select high-impact articles from diverse sources, including journals, dissertations, monographs, and ERIC documents, ensuring the relevance and rigor of the selected studies.

### Research Design

The study follows a thematic analysis design, which is suited for identifying and interpreting patterns across a dataset. This design facilitates a comprehensive exploration of how error perception is conceptualized and evaluated in both spoken and written contexts, as well as the criteria used to establish hierarchies of error types.

### Population and Sampling

The study targets a population of original articles in the field of error gravity, emphasizing works that examine error perception and evaluation. Using purposive sampling, 31 articles were chosen based on their relevance, citation impact, and contribution to understanding error gravity.

### Data Collection and Analysis

I systematically extracted key components from the selected articles, including research objectives, participant demographics, methodologies, and findings, as part of the data collection process. Following this, I generated and finalized four main themes: *Perceptual Dichotomy: Leniency vs. Seriousness of Errors* (exploring the continuum of error severity perceptions), *Error Perception in Spoken Communication* (understanding how oral errors are evaluated in different communicative contexts), *Error Evaluation in Written Composition* (analyzing evaluative criteria applied to errors in writing), and *Establishing Hierarchies of Error* (investigating how errors are ranked by severity and impact).

### Methodological Framework

The study applies a qualitative synthesis framework based on thematic analysis. This framework supports a systematic examination of patterns within the literature, enabling a comprehensive discussion of the identified themes. The methodology ensures alignment with the research objectives, providing a robust basis for the thematic analysis presented in the Results and Discussion section.

## Results and Discussion

This research reviews numerous high-impact studies on error gravity, ranging from Bansal's (1969) work to Porte's (1999) study. As this study focuses on reviewing these articles from the point of view of leniency in error analysis, perceptions of errors in spoken and written language, and the hierarchy of error types, I have discussed the findings under the following headings.

### Perceptual Dichotomy: Leniency vs. Seriousness of Errors

James (1977), Hughes and Lascaratou (1982), Davies (1983), Majer (1985), Sheorey (1986), and McCretton and Rider (1993) are renowned authors who, through their high-impact articles, have significantly contributed to the field of applied linguistics for many years. Their work has become a landmark in understanding error gravity, balancing leniency with a deeper focus on the seriousness of linguistic errors. These contributions have made history in applied linguistics and will continue to inspire future scholars. These authors are frequently cited and widely referenced by students and teachers, and research scholars around the world. Their articles have been reviewed countless times and have become iconic resources in the study of error gravity, benefiting students of applied linguistics globally.

Research studies have reached a unanimous conclusion that native speakers of English demonstrate leniency toward ESL errors, while EFL learners tend to evaluate these errors more harshly. This analysis ranges from James's (1977) "Judgements of Error Gravities" to Porte's (1999) "Where to Draw the Red Line: Error Toleration of Native and Non-native EFL Faculty." By examining the objectives and findings of these studies, a clear trend emerges: native speakers generally adopt a more tolerant approach to errors, which reflects their familiarity with the language, whereas EFL learners are more critical in their evaluations. This difference shows that factors like language background and teaching environment strongly affect how errors are viewed. Overall, these studies highlight the complex nature of error evaluation in language learning, emphasizing the need to consider the context of learners in educational settings.

James (1977) aims to evaluate the seriousness of errors made by learners of English as a Foreign Language (EFL) from the perspective of both native and non-native EFL teachers. James (1977) concludes that non-native teachers (nationals) tend to judge the errors more harshly than native English speakers. Specifically, native speakers rated errors at 49.2% on average, while non-native teachers rated them at 55.2%. This finding confirms Nickel's observation (Nickel, 1970, p. 10) that native speakers are probably more tolerant of learners' errors than teachers of the same nationality as the students. This finding could have implications for language teaching practices and the attitudes of

educators towards student errors. Hughes & Lascaratou (1982) study also aims to evaluate the seriousness of English language errors made by Greek-speaking students which are judged by 10 native-speaker teachers of English, 10 educated English native-speakers who are not teachers and 10 Greek teachers of English. The results revealed that both groups of native-speaker teachers and non-teachers were more lenient in judging the seriousness of errors than the Greek teachers. The sentence ratings of the Greek teachers correlated with those of the English teachers at 0.40 and with English non-teachers at only 0.07, while English teachers' and non-teachers' ratings correlated at 0.82 (p. 178).

Davies (1983) aims to examine how teachers' experience and knowledge of their pupils influence their judgments of error gravity, unlike James (1977) and Hughes and Lascaratou (1982), who aimed to evaluate the seriousness of ESL error. The study compares these judgments with those of non-teachers and investigates differences between native and non-native speakers. The results provide further evidence that native judges, despite not being teachers, are more lenient in their evaluations compared to non-native judges who are English teachers. Native speakers tend to focus on overall communication rather than technical correctness, which contributes to their leniency. In contrast, non-native teachers often adopt a stricter stance, likely due to their professional training and greater awareness of linguistic rules and norms. In this study, the sample population were 43 Moroccan natives, who were teachers of English, and 43 native English speakers from Britain, who were non-teachers. On average, native speakers assigned scores which were on average approximately one point lower than those of the non-native speakers. There may be more than one reason for this contrast. Hughes and Lascaratou (1982) follow Nickel (1973) in suggesting that the native speakers' tendency to greater leniency may be attributed to their superior knowledge of the wide ranging norms of English.

Majer (1985) examines the evaluations of ESL errors by Polish teachers, American teachers, and non-teachers, comparing their error grading patterns. The present study provides further evidence of the tendency of native speakers to evaluate learner errors more leniently than non-native speakers. The Polish judges' scores of 3.35, American teachers' scores of 2.54, and American non-teachers' scores of 2.34 indicate that Polish teachers assigned grades that were, on average, one point higher than those assigned by American non-teachers and 0.8 points higher than those assigned by American teachers. Overall, these findings align with those reported by James (1977), Hughes and Lascaratou (1982), and Davies (1983), which reinforce the trend of leniency among native speakers in error evaluation.

Sheorey (1986) compares the error perceptions of 62 American native-speaking ESL teachers and 34 Indian non-native ESL teachers, aligning with Majer's (1985) study on

error grading pattern. Sheorey (1986) examined whether consistent and statistically significant differences existed in error gravity perceptions between native and non-native ESL teachers, finding that native teachers were more tolerant of errors. The results indicated that there was a difference of slightly over nine points between the average scores of the two groups, with non-native teachers deducting an average of 59.82 points out of a possible 100, compared to native teachers, who deducted an average of 50.19 points. This difference was statistically significant ($p < .01$) (Sheorey, 1986, p. 308). This finding is consistent with the findings of James (1977) and Hughes and Lascaratou (1982) in that native speakers (teachers as well as non-teachers) appear to be more tolerant of errors made by ESL students than the native speakers are.

McCretton and Rider (1993) aimed to judge the gravity of written errors by having 10 native English-speaking teachers and 10 non-native English-speaking teachers who evaluate 25 sentences containing seven types of errors. The results showed that non-native speaker teacher (NNST) judges were consistently more severe in their judgments than native speaker teacher (NST) judges. The author concludes that error hierarchies are neither inherent nor universal but instead reflect the judges' educational training. In attempts to explain this result, the authors suggest that non-native speaking teachers "may have felt that their own knowledge of the language was being tested, and as a reaction to this, tended to mark more severely" (p. 182).

Above all we find that James (1977), Hughes and Lascaratou (1982), and Davies (1983) focused on assessing the seriousness of errors, but the profiles of the native speakers evaluating these errors varied. In James's (1977) study, the evaluators were native speaker teachers of English who were not nationalized. Hughes and Lascaratou (1982) involved both teachers and non-teachers, neither of whom was nationalized. Meanwhile, Davies's (1983) study included native English speakers from Britain who were non-teachers. Similarly, Majer (1985) examines the evaluation of ESL errors by Polish teachers, American teachers, and non-teachers, comparing their error grading patterns. Sheorey (1986) also explores the error perceptions of native-speaking American teachers and non-native ESL teachers. McCretton and Rider (1993) focused on assessing the gravity of written errors as judged by native English-speaking and non-native English-speaking teachers. Overall, the findings reveal that native speakers, regardless of their country and professional experience, tend to perceive errors more leniently than non-native speakers, whether they are Indian, Polish, or Greek. In contrast, non-native speakers consistently evaluate ESL errors more harshly.

This review study also revealed that the number of respondents in these studies shows no correlation with the findings, and none of the researchers addressed how population size influenced their results. For example, whether the population consisted of 20 participants

(10 native English teachers and 10 non-native teachers, as in McCretton and Rider, 1993), 30 participants (20 native speakers and 10 non-native teachers, as in Hughes and Lascaratou, 1982), 40 participants (20 native speakers and 20 nationals, as in James, 1977) or 96 participants (62 American native speaker teachers and 34 Indian non-native teachers, as in Sheorey, 1986), the findings remained unaffected by the sample size. This highlights a critical inconsistency in the research: the selection of varying numbers of evaluators in error gravity studies lacks justification, as these variations do not influence the outcomes.

**Error Perception in Spoken Communication**

Several research articles I reviewed relating to error perceptions in spoken English include Bansal (1969), Politzer (1978), Albrechtsen, Henriksen, and Færch (1980), Fayer and Krasinski (1987), and Hadden (1991). The study of error perception in spoken language primarily focused on data produced by non-native English speakers from diverse linguistic backgrounds, including India, Germany, Denmark, and China. Similarly, the judges in these studies included both native and non-native English speakers, such as individuals from Britain, the United States, Nigeria, and Germany, challenging the expectation that evaluators must be native English-speaking teachers or educators. Unquestionably, the data under evaluation exclusively belonged to non-native speakers, forming the basis of the present theme.

Several studies have examined how errors in learners' spoken English are perceived by both native and non-native speakers. Among these, Bansal's (1969) research on the intelligibility of spoken English inherently deals with the perceptions of pronunciation-related errors. His study specifically analyzed speech recordings of 24 Indian English speakers to evaluate how intelligible Indian English was to listeners from various linguistic backgrounds. The evaluators from the UK, US, Nigeria, and Germany offered valuable insights into the comprehension of Indian English by both native and non-native speakers. The study revealed that certain phonetic peculiarities of Indian English hinder intelligibility, as revealed through listener evaluations.

Unlike Indian English which served as the corpus for Bansal's (1969) study, Politzer (1978) examines errors made by English speakers learning German, focusing on how German native speakers perceive and evaluate these errors. German native speakers as the judges evaluated the recording of sixty (60) pairs of German sentences, all containing deviations from standard German. Politzer (p. 259) observes that vocabulary errors are considered more serious than grammar errors, with non-native pronunciation deemed the least important among the three factors evaluated. This contrasts with the findings of Ensz (1982) who reported that native French speakers rated grammar errors as most important, followed by vocabulary and then pronunciation errors in English-accented

French. These variations may reflect cultural or language-specific perceptions of error severity among native speakers of German (in Politzer's (1978) study) and French (in Ensz's (1982) study).

Other studies have been conducted on native English speakers' reactions to spoken English. Both Albrechtsen, Henriksen, and Færch (1980) and Fayer and Krasinski (1987) studied native English speakers' reactions to spoken language. The former focused on stretches of discourse during interviews by Danish learners, aiming to assess how native speakers perceive spoken language, which are typically studied in written performance. In contrast, Fayer and Krasinski (1987) examined reactions to tape recordings of Puerto Rican learners of English at various proficiency levels, comparing the responses of native English and Spanish speakers.

Albrechtsen, Henriksen, and Færch's (1980) study involved 150 British native speakers who rated learner speech using 14 five-point bipolar adjective scales. Factor analysis identified four main factors: personality, content, language, and comprehension. The results showed a strong correlation between the language factor and most performance features, while the comprehension factor showed only one significant correlation—communication strategies. In Fayer and Krasinski's (1987) study, listeners completed questionnaires evaluating variables such as intelligibility, grammar, pronunciation, intonation, vocabulary, hesitations, and distraction. The results revealed differences between English and Spanish listeners, with Spanish listeners showing less tolerance for non-native speech. Despite this, both groups identified pronunciation and hesitations as the most distracting features affecting message clarity.

Hadden (1991) compares ESL teachers' (n=25) and non-teachers' (n=32) perceptions of second-language communication. This study involved eight native speakers of Chinese speaking extemporaneously in videotapes, which were evaluated by 57 native speakers of American English through a questionnaire. The study found that perceptions were complex and multifaceted, not simple. A factor analysis revealed that while the two groups differed slightly in their identified factors, the combined data reflected five consistent factors: comprehensibility, social acceptability, linguistic ability, personality, and body language.

Here is a summary of the results on the basis of the above review studies:

Bansal's (1969) study highlighted the significant impact of phonetic peculiarities in Indian English on intelligibility, as evidenced by listener evaluations. These findings emphasized the critical role of pronunciation in effective spoken communication. In contrast, Politzer's (1978) research identified vocabulary errors as the most serious, followed by grammar errors, with pronunciation errors ranked as the least significant.

This hierarchy of error severity diverged from Ensz's (1982) findings on French, underscoring the influence of cultural and linguistic contexts on error perception.

Albrechtsen, Henriksen, and Færch (1980) further clarified the complexity of error evaluation in spoken discourse. Their study demonstrated a strong correlation between the language factor and various performance features, while comprehension was closely linked to communication strategies, reflecting the multifaceted nature of spoken language assessment. Similarly, Fayer and Krasinski (1987) revealed differences in tolerance levels between Spanish and English listeners, with Spanish listeners being less forgiving of non-native speech. Nevertheless, both groups agreed that pronunciation and hesitations were the most distracting features, significantly affecting message clarity. Hadden's (1991) findings reinforced this complexity, identifying five consistent factors—comprehensibility, social acceptability, linguistic ability, personality, and body language—as central to perceptions of second-language communication.

**Error Evaluation in Written Composition**

The perception of errors in written language has been the subject of several notable studies, each examining distinct aspects of this complex phenomenon. Researchers such as Delisle (1982), Vann, Meyer, and Frederick (1984), Santos (1988), Bader (1988), Kobayashi (1992), and Porte (1999) have focused on understanding how errors in written language are evaluated. These studies have explored diverse areas, including the development of error hierarchies, the assessment of error severity, and the analysis of university professors' responses to written compositions and academic writing.

Delisle (1982) conducted a study on error perception in written language, involving 193 high school students aged 13 to 17, with just over half being girls, who attended a Gymnasium in Hamburg, Germany. The study aimed to establish the role of medium—spoken or written—in the evaluation of errors. Delisle's research is a modified version of Politzer's (1978) study, which focused on error perception in spoken language. In Delisle's (1982) study, a hierarchy of errors was established based on their perceived severity by native speakers. The students identified gender errors (63%) in L2 German as the most serious. This finding, however, does not align with Politzer's (1978) study, in which 146 German native-speaking teenagers identified vocabulary errors as the most serious.

Vann, Meyer and Frederick (1984) focus on sentence-level errors in which the authors attempt to determine which sentence-level errors are judged to be most serious by an academic community and then to discover what factors may influence this judgment. A survey was conducted to measure how a cross-section of faculty at Iowa State University responds to certain common ESL writing errors. There were 164 respondents who participated in this research who completed a 3-page questionnaire and ranked the

relative gravity of 12 typical ESL written errors occurring in 24 sentences. Results indicate that most respondents did not judge all errors as equally grievous; rather, their judgements generate a hierarchy of errors. The study also suggests that both the age and academic discipline of faculty members may be important factors in predicting their response to certain ESL student writing errors.

Santos (1988) aims to investigate the reactions of university professors to two 400-word compositions written by a Chinese and a Korean student. A total of 178 professors representing different academic disciplines participated in the study, which elicited judgments about the content and language of the compositions. The results were as follows: (a) Content received lower ratings than language; (b) professors found the errors highly comprehensible, generally unirritating, but academically unacceptable, with lexical errors rated as the most serious; (c) professors in the humanities/social sciences were more lenient in their judgments than professors in the physical sciences; (d) older professors were less irritated by errors than younger professors, and non-native-speaking professors were more severe in their judgments than native speakers. Non-native speaking professors "have attained an extremely high level of proficiency in English and, because of their investment of effort in the language, judge the errors of other NNSs [non-native speakers] more severely than do NS professors" (Santos 1988, p. 85).

There are also effects of age and specialization. Professors in the physical sciences rated the language quality of the compositions as significantly less acceptable compared to their counterparts in the humanities and social sciences (Santos, 1988, p. 80). This result was confirmed by Roberts and Cimasco (2008) who additionally found that "the older professors displayed a lower degree of irritation in their ratings than did the younger professors" (p. 81). The results were the same for both compositions: The content was rated significantly lower than the language. The professors judged the content more severely than the language.

Kobayashi's (1992) study involves 269 judges, including English native speakers (ENSs) and Japanese native speakers (JNSs), at the professorial, graduate, and undergraduate levels. These judges evaluate and edit two ESL compositions written by Japanese college students, using 10-point scales to assess grammaticality, clarity of meaning, naturalness, and organization. The findings indicate that ENSs are stricter in grammatical assessments and make significantly more corrections with higher accuracy compared to JNSs. Both groups show that higher academic status correlates with greater accuracy in error correction. ENS professors and graduate students provide more positive evaluations for clarity of meaning and organization compared to their Japanese-speaking counterparts. JNSs, however, often leave certain errors uncorrected, particularly those related to articles, number, prepositions, and loanword-based lexical items.

Porte (1999) aims to investigate professors' reactions to the academic writing of non-native students. Fourteen native (14) and sixteen non-native (16) university professors responded the student errors based on a random selection of fifty-four anonymous ungraded compositions written by students at Granada University. General comparison of the error grading between the two groups of subjects revealed small differences in the error toleration of native and non-native faculty. While differences do exist in the perceived gravity of specific errors, it would appear that teachers in this study generally agreed in their judgments. Nevertheless, there was evidence that errors are not being perceived as seriously as one would have expected and the implications of this finding are discussed.

Here is a summary of the findings based on the reviewed studies on error evaluation in written composition.

Delisle (1982) established a hierarchy of errors in written language, finding that gender errors were perceived as the most serious by 193 high school students in Hamburg, Germany. This finding contrasts with Politzer's (1978) study, which identified vocabulary errors as the most serious among German teenagers. Vann, Meyer, and Frederick (1984) examined sentence-level errors, revealing that academic faculty at Iowa State University ranked ESL writing errors hierarchically rather than equally, influenced by factors such as faculty age and academic discipline. Santos (1988) explored university professors' reactions to compositions by non-native speakers, concluding that while content was rated lower than language quality, lexical errors were considered the most serious. Notably, professors in the humanities and social sciences were more lenient than their counterparts in the physical sciences, and older professors exhibited less irritation towards errors than younger ones. Additionally, non-native-speaking professors tended to judge errors more harshly than native speakers. Kobayashi (1992) findings indicated that English native speakers were stricter in grammatical assessments and made more corrections than their Japanese counterparts, with higher academic status correlating with greater accuracy in error correction. Porte (1999) investigated professors' reactions to non-native students' academic writing, revealing small differences in error tolerance between native and non-native faculty. While some discrepancies existed in the perceived severity of specific errors, overall, both groups showed agreement in their evaluations, suggesting that errors may not be perceived as severely as expected.

**Establishing Hierarchies of Error Types**

This review study seeks to establish hierarchy of errors committed by ESL learners at various levels. The pioneering researchers like James (1977) to McCretton and Rider (1993) worked on ranking the error types.

James (1977) revealed a hierarchy of grammatical errors, ranked from most to least severe as transformations, tense, concord, case, negation, articles, word order, and lexical errors, based on evaluations by native and non-native speakers of English. Politzer's (1978) study aims to determine the relative importance that native speakers assign to different types of errors—specifically vocabulary, grammar, and pronunciation errors—in spoken German. Politzer (p. 259) concludes that vocabulary errors are regarded as more serious than grammar errors, while non-native pronunciation is considered the least important of the three factors evaluated. Some years later, Delisle (1982) repeated the investigation making adjustments for the written language. Delisle (1982) offers a hierarchy of error types based on the evaluations by German 13-17 year-olds of written sentences produced by English speakers learning German. The German students considered gender error the most serious, followed by verb morphology and spelling error as the least serious categoty. In both studies, the subjects evaluating the errors were students in West German schools. These students assessed sixty pairs of sentences, each containing an error. However, the studies differed in their error hierarchies.

Based on the responses of native-speaking English teachers, Hughes and Lascaratou (1982, p. 178) established an error gravity hierarchy. In this hierarchy, pronoun and vocabulary errors are regarded as the most serious, while preposition and concord (subject-verb agreement) errors are considered less serious. Overall, these grammatical errors rank from most serious to least serious as follows: pronouns, vocabulary, verb forms, plurals, spelling, word order, concord, and prepositions. Vann, Meyer, and Frederick (1984) developed a hierarchy of errors, identifying word order errors as the least acceptable with a mean ranking of 1.8, while spelling errors were judged the most acceptable, with a mean value of 3.2. This hierarchy aligns with intuitive perceptions regarding the seriousness of different errors and supports Burt and Kiparsky's (1975) findings on the adverse communicative impact of errors, particularly in areas like word order that affect overall sentence structure. Moreover, it suggests that native speakers are generally less tolerant of errors perceived as more 'foreign'. We find similar findings discussed in Sheorey's (1986) study, where Sheorey (1986) suggests that we should be more tolerant overall in evaluating error, give lower priority to spelling errors and place greater importance on lexical nuances than we normally do. Although both studies emphasize the importance of communicative effectiveness, Sheorey’s (1986) study differs by suggesting that spelling errors should not be a primary concern, in contrast to Vann, Meyer, and Frederick's (1984) conclusion that spelling errors are the most acceptable type of error. Santos (1988) highlighted the utmost gravity of lexical errors in her study, where they were judged to be the most serious error type in student compositions. McCretton and Rider (1993) proposed a hierarchy of errors in ESL based on seriousness, starting with lexis as the most serious, followed by spelling, negation,

word order, prepositions, verb forms, and concord as the least serious. Nearly every error evaluation study reviewed here provided some evidence to support the assumption that some errors are considered more serious than others. However, evidence for a "universal hierarchy of errors" is inconclusive. McCretton and Rider (1993) concluded that error hierarchies are not inherent and "universal" but reflect the evaluators' own educational training.

## Conclusion

Error gravity was a prominent and widely discussed topic in the 1970s and 1980s but seemed to fall out of favor in the 1990s. Resonating with this trajectory, Endley (2016) notes that only recently have researchers begun to renew their interest in this issue. Among the error gravity studies, various variables—such as comprehensibility, intelligibility, acceptability, irritation, and grammaticality—were employed in evaluating ESL errors. These evaluations often involved judges from diverse backgrounds, including native and non-native English-speaking teachers, as well as non-teachers, with professors participating on a large scale. Furthermore, evaluators' age, gender, profession, and educational background were identified as factors influencing their perceptions of errors. Likert scales were commonly used to rate errors from most to least serious. These methodological practices, documented in the error gravity articles under review, continue to shape how errors are evaluated by students, teachers, and non-teachers across various contexts and time periods.

Some errors in language learning are considered more serious, while others are classified as less significant. Evaluators often face a dilemma in determining which errors warrant the most attention and require special treatment. However, there is no universally accepted framework or conclusive modality for identifying and ranking errors with high gravity. Although many researchers have proposed guidelines for evaluating ESL errors (James, 1998), these frameworks have not been entirely effective in practice. Without clear guidelines, evaluators usually rely on their own criteria, which they develop based on their experience and understanding. This means the way errors are judged can vary a lot depending on who is doing the evaluation. Establishing standard guidelines is challenging because different contexts and individual perspectives play a significant role in deciding which errors matter more.

This study highlights evaluation of ESL errors in spoken communication and written compositions differently, with evaluators ranking grammatical errors variably. Native German speakers evaluated errors in both modalities. For spoken communication, they considered vocabulary errors more serious than grammar errors (Politzer, 1978). In written communication, however, Delisle (1982) observed that the native German speaking students identified gender errors in L2 German (63%) as the most serious. This

proves the hierarchy of error gravity across studies reveals varying patterns. Pronouns (Hughes and Lascaratou, 1982), word order (Vann, Meyer, and Frederick, 1984), transformations (James, 1977), and concord (McCretton and Rider, 1993) have been identified as critical areas of concern. However, evidence for a "universal hierarchy of errors" remains inconclusive, and no consistent pattern has emerged regarding how students perceive and apply error hierarchies in their studies. Despite these variations, the findings consistently reflect that contextual factors and the evaluators—whether native or non-native English speakers—play a significant role in shaping the ranking of errors. The evaluators' educational background has been found to play a crucial role in shaping error hierarchies.

## References

Albrechtsen, D., Henriksen, B., & Færch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning, 30*(2), 365–396. https://doi.org/10.1111/j.1467 -1770.1980.tb00324.x

Bader, F. M. N. (1988). *Analysis of error type, source, and gravity in the writing of Arabic ESL students in U.S.A. colleges* (Publication No. 3753) [Master of Arts dissertation, Portland State University]. Dissertations and Theses. https://doi.org/10.15760/etd.5637

Bansal, R. K. (1969). *The intelligibility of Indian English* (Monograph No. 4). Central Institute of English and Foreign Languages.

Burt, M., & Kiparsky, C. (1975). Global and local mistakes. In J. Schumann & N. Stenson (Eds.), *New frontiers in second language learning*. Rowley, MA: Newbury House.

Chastain, K. (1980). Native speaker reaction to instructor-identified student second-language errors. *The Modern Language Journal*, *64*(2), 210-215.

Davies, E. E. (1983). Error evaluation: The importance of viewpoint. *ELT Journal, 37*(4), 304–311. https://doi.org/10.1093/elt/37.4.304

Delisle, H. H. (1982). Native speaker judgement and the evaluation of errors in German. *The Modern Language Journal, 66*(1), 39-48. https://doi.org/10.1111/j.1540-4781.1982.tb01020.x

Endley, M. (2016). Perceptions of error gravity among native speaking and non-native speaking. In *Proceedings of the International Conference on Applied Linguistics* (pp. 123–134). Cambridge Scholars Publishing.

Ensz, K. (1982). French attitudes toward typical speech errors of American speakers of French. *The Modern Language Journal, 66*(2), 133–139.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*(3), 313–326. https://doi.org/10.1111/j.1467-1770.1987.tb00573.x

Guntermann, G. (1978). A study of the frequency and communicative effects of errors in Spanish. *The Modern Language Journal, 62*(5/6), 249-253. https://doi.org/10.2307/324906

Hadden, B. L. (1991). Teacher and non-teacher perceptions of second language communication. *Language Learning, 41*(1), 1–24. https://doi.org/10.1111/j.1467-1770.1991.tb00674.x

Hughes, A., & Lascaratou, C. (1982). Competing criteria for error gravity. *ELT Journal, 36*(3), 175–182. https://doi.org/10.1093/elt/36.3.175

Hultfors, P. (1986). *Reactions to non-native English: Native English speakers' assessments of errors in the use of English made by non-native users of the language. Part 1: Acceptability and intelligibility* [Doctoral dissertation, Stockholm University, Faculty of Humanities, Department of English]. Stockholm University.

Ilin, N. (2017). *Perception of learner errors and non-standard features in the native and non-native language: Evaluation vs. processing cost* (Doctoral thesis). Albert-Ludwigs-Universität Freiburg.

James, C. (1972). The diagnosis and analysis of errors: Some insights from linguistics. *Audio-Visual Language Journal, 10*, 75–79.

James, C. (1977). Judgements of error gravities. *31*(2), 116–124. https://doi.org/10.1093

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. Longman.

Johansson, S. (1978). *Studies of error gravity: Native reactions to errors produced by Swedish learners of English*. Goteborg, Sweden: Acta Universitatis Gothoburgensis.

Khalil, A. (1985). Communicative error evaluation: Native speakers' evaluation and interpretation of written errors of Arab EFL learners. *TESOL Quarterly, 19*(2), 335-351.

Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly, 26*(1), 81–112. https://doi.org/10.2307/3587370

Ludwig, J. (1982). Native-speaker judgments of second-language learners' efforts at communication: A review. *The Modern Language Journal, 66*(3), 274–283. https://doi.org/10.2307/326629

Majer, J. (1985). Native vs. non-native judgments of error gravity: An experimental study. *Studia Filologiczne (Philological Studies), 21,* 91–107. http://repozytorium.ukw.edu.pl/handle/item/2593

McCretton, E., & Rider, N. (1993). Error gravity and error hierarchies. *IRAL - International Review of Applied Linguistics in Language Teaching, 31*(3), 177–188. https://eric.ed.gov/?id=EJ475974

Nickel, G. (1970). Zum Problem der Fehlerbewertung. In *PAKS-Aibeishericht Nr. 5*. Stuttgart, Germany.

Oliaei, S., & Sahragard, R. (2013). Error treatment by experienced and inexperienced Iranian EFL teachers of writing. *Educational Research International*, *2*(1), 55-75.

Olsson, M. (1972). *Intelligibility*. The GUME Project, No. 12. Department of Educational Research, Gothenburg School of Education.

Olsson, M. (1973). The effects of different types of errors in the communication situation. *In J. Svartvik (Ed.), Errata* (pp. 153–160). Lund: Gleerup.

Piazza, L. G. (1980). French tolerance for grammatical enors made by Americans. *The Modern Language Journal*, *64*(4), 422-427.

Politzer, R. L. (1978). Errors of English speakers of German as perceived and evaluated by German natives. *The Modern Language Journal, 5*(6), 253 – 61. DOI:10.1111/J.1540-4781.1978.TB02395.X

Porte, G. (1999). Where to draw the red line: Error toleration of native and non-native EFL faculty. *Foreign Language Annals*, *32*(4), 426-436. https://doi.org/10.1111/j.1944-9720.1999.tb00873.x

Rifkin, B., & Roberts, F. (1995). Error gravity: A critical review of research design. *Language Learning*, *45*(3), 511-537. http://dx.doi.org/10.1111/j.1467-1770.1995.tb00450.x

Rifkin, B., & Roberts, F. D. (1995). Error gravity: A critical review of research design. *Language Learning, 45*(3), 511-537. https://doi.org/10.1111/j.1467-1770.1995.tb00450.x

Roberts, F., & Cimasko, T. (2008). Evaluating ESL: Making sense of university professors' responses to second language writing. *Journal of Second Language Writing, 17*(3), 125–143. https://doi.org/10.1016/j.jslw.2007.10.002

Robinson, P. (1973). Testing the second-language competence of children and adults. *English Language Teaching, 27*, 190–199.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*(1), 69–90. https://doi.org/10.2307/3587062

Schmitt, N. (1993). Comparing native and nonnative teachers' evaluations of error seriousness. *Japanese Association of Language Teachers Journal*, *15*(2), 181-191.

Sheorey, R. (1986). Error perceptions of native-speaking and non-native-speaking teachers of ESL. *ELT Journal*, *40*(4), 306–312. https://doi.org/10.1093/elt/40.4.306

Takashima, H. (1987). To what extent are non-native speakers qualified to correct free composition? A case study. *The British Journal of Language Teaching*, *25*(1), 43-48.

Tong, A. K. K. (2000). *A phenomenological study of ESL instructors' perceptions of written errors among adult ESL learners* (Doctoral dissertation). University of Toronto.

Vann, R. J., Meyer, D. E., & Frederick, L. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly, 18*(3), 427-440. https://doi.org/10.2307/3586713