

Digital Code-Switching and Language Attitudes: A Mixed-Methods Study of English–Sindhi–Urdu Code-Mixing in WhatsApp Communication among University Students in Sindh, Pakistan

Dr. Ali Siddiqui

Lecturer- English, Hyderabad Institute for Technology and Management Sciences (HITMS),
drali@faculty.hitms.edu.pk

Razia Hataf

M.Phil. Scholar, Center of English Language and Linguistics,
Mehran University of Engineering and Technology, Jamshoro, Sindh, Pakistan
razia.hataf@outlook.com

Shamshad Junejo

Lecturer-English, Center of English Language and Linguistics,
Mehran University of Engineering and Technology, Jamshoro, Sindh, Pakistan
Shamshad.junejo@faculty.muett.edu.pk

Abstract

Mobile-mediated communication has reshaped multilingual practices in Pakistan, yet digital interaction among Sindhi university students remains under-researched. This study examines the frequencies, drivers, and attitudes regarding Sindhi–Urdu–English code-switching on WhatsApp among 280 postsecondary students in Sindh. Framed by Myers-Scotton’s Markedness Model, this convergent mixed-methods inquiry evaluated survey responses using SPSS (v26), alongside a 42,150-word chat corpus and 15 semi-structured interviews via NVivo (v14). Quantitative findings reveal Sindhi–English as the most prevalent combination (34.6%), followed by Urdu–English (29.3%) and trilingual integration (27.5%). Gender patterns show female participants switching predominantly for relational and identity functions, whereas male peers favor technical and topical switches. Overall, respondents expressed favorable attitudes, highlighting ease of communication and vocabulary bridging. Qualitative thematic synthesis yielded six core motivations: convenience, identity construction, peer solidarity, status, lexical gaps, and humor. The study contends that WhatsApp functions as a distinct sociolinguistic domain where youth perform a dynamic, technology-mediated identity that both resists and reinforces English-dominated linguistic hierarchies. To better explain digital language selection, the author proposes "digilingual markedness" as an extended adaptation of the Markedness Model. These insights provide significant contributions to sociolinguistic theory, educational policy, and English language pedagogy in Sindh.

Keywords: Digital Code-Switching, Multilingualism, WhatsApp Communication, Digilingual Markedness

1. Introduction

Pakistan is a multilingual state in which more than seventy languages coexist alongside Urdu, the national language, and English, the de facto official language of governance, higher education, and elite social mobility (Rahman, 2002; Mahboob, 2009). Sindh, Pakistan's second-most populous province, presents a particularly rich sociolinguistic ecology in which Sindhi functions as the primary mother tongue and marker of ethnic identity, Urdu operates as the lingua franca of inter-ethnic communication, and English continues to index modernity, prestige, and access to economic opportunity (Channa, Shah, Manan, & David, 2025; Kaloi, Sagheer, & Khalil, 2024). Historically, this trilingual configuration has been studied largely through the lens of classroom code-switching, language policy, and language shift (Manan, David, & Dumanig, 2015, 2017; David, Ali, & Baloch, 2017). However, the rapid expansion of smartphone ownership and social media use in Pakistan over the past decade has generated a new and largely under-examined domain of language contact: informal, text-based, peer-to-peer digital communication.

This shift is not merely technological but deeply sociolinguistic. Whereas earlier generations of Sindhi speakers acquired and deployed English primarily through formal schooling, print media, and broadcast television (Coleman, 2010; Mansoor, 2005), the present generation of university students encounters and produces English in a continuous, informal, peer-regulated stream through instant messaging applications. This changes not only the quantity of English exposure but also its qualitative character: English on WhatsApp is typically informal, abbreviated, emoji-accompanied, and produced under conversational rather than institutional norms, raising the question of whether theoretical models developed for spoken or broadcast code-switching remain adequate for describing this new digital register (Crystal, 2003).

According to DataReportal's (2024) Digital 2024: Pakistan report, Pakistan had 111.0 million internet users and 71.70 million active social media identities at the start of 2024, while the Pakistan Telecommunication Authority (2024) reported that social media accounts in the country had crossed 303 million profiles by mid-2024, with WhatsApp remaining the dominant platform for personal and peer communication among young adults. For university students in Sindh, WhatsApp has become the primary channel through which academic, social, and familial relationships are sustained, and it is within this digital space that Sindhi, Urdu, and English are in constant, often unconscious, contact. Existing Pakistani research on digital code-switching has focused predominantly on Punjabi–English or Urdu–English patterns in Facebook and WhatsApp interactions (see, e.g., studies reviewed in Section 2), leaving the specifically trilingual, Sindhi-inflected pattern of digital code-switching comparatively unexplored.

This gap is theoretically significant because Sindhi occupies a distinctive sociolinguistic position in Pakistan: unlike several other indigenous languages, it possesses official provincial status, a standardised script, and a long literary tradition (Channa et al., 2025; David et al., 2017), yet it is simultaneously subject to lexical erosion through sustained contact with English and Urdu (Buglio, 2006; Abbasi, 2023). Understanding how young, educated Sindhi speakers navigate this trilingual repertoire in the informal, low-stakes environment of WhatsApp chat offers a valuable window onto broader processes of language change, identity negotiation, and attitude formation among Pakistan's digital generation.

The present study therefore examines three interrelated questions: (1) What patterns and frequencies of code-switching and code-mixing occur in the WhatsApp communication of university students in Sindh? (2) What attitudes do these students hold toward code-switching, and do these attitudes vary by gender? (3) What communicative functions and social meanings do students themselves attribute to their switching behaviour? To answer these questions, the study adopts a convergent mixed-methods design combining a quantitative questionnaire survey with qualitative corpus and interview data, framed within Myers-Scotton's (1993) Markedness Model of code-switching. In doing so, the study extends the Markedness Model into the domain of computer-mediated communication (CMC) and proposes the concept of digilingual markedness to capture how markedness is renegotiated when a switch occurs in text rather than speech.

The significance of this inquiry is threefold. Theoretically, it extends a well-established sociolinguistic framework into an increasingly dominant but under-theorised communicative domain, offering a model that may be transferable to other multilingual, digitally connected youth populations across South Asia. Empirically, it provides one of the first systematically triangulated datasets on trilingual Sindhi–Urdu–English code-switching, addressing a population that has been comparatively neglected relative to Punjabi- and Pashto-speaking communities in existing Pakistani CS literature. Practically, the findings carry implications for English language teaching, curriculum design, and language policy debates in Sindh, where questions of mother-tongue-based multilingual education continue to be actively contested (Manan et al., 2017).

2. Review of Literature

This section reviews the theoretical and empirical literature on code-switching, language attitudes, and digital multilingualism, with particular attention to the Pakistani and Sindhi contexts, before presenting the theoretical framework that guides the present study.

2.1 Code-Switching: Definitions and Theoretical Traditions

Code-switching (CS) refers to the alternating use of two or more linguistic codes within a single conversation, utterance, or even clause (Gumperz, 1982; Poplack, 1980). Blom and Gumperz's

(1972) foundational study distinguished between situational code-switching, triggered by a change in setting, participants, or topic, and metaphorical code-switching, in which a switch communicates a shift in social meaning without a corresponding change in the external situation. Auer (1998) further developed a sequential, conversation-analytic approach to CS, emphasising that the meaning of a switch is negotiated turn-by-turn between interlocutors rather than fixed in advance. Poplack (1980) additionally distinguished structural types of switching, namely inter-sentential switching (between clauses or sentences), intra-sentential switching (within a single clause), and tag-switching (insertion of short fixed expressions), a typology that remains widely used in contemporary corpus-based CS research, including studies of Pakistani digital communication (Blending Words, Bridging Cultures, 2025).

Fishman's (1972) domains-of-language-use framework offers a complementary lens, proposing that speakers in multilingual societies allocate different languages to different social domains, such as family, education, religion, and government. In Pakistan, this domain allocation has historically placed English in the domains of education, bureaucracy, and elite social interaction, Urdu in domains of national and inter-ethnic communication, and indigenous languages such as Sindhi in domains of home, community, and ethnic solidarity (Rahman, 2002; Mansoor, 2005). The rise of WhatsApp as an informal, largely unregulated communicative space arguably destabilises this traditional domain allocation, allowing all three codes to intermingle within a single, previously home- or peer-oriented domain.

2.2 Code-Switching and Language Attitudes in Pakistan

A substantial body of Pakistani sociolinguistic research documents positive student attitudes toward classroom code-switching between English and Urdu or English and Sindhi, typically framing CS as a pedagogical facilitation strategy rather than a linguistic deficiency (Talpur et al., 2021, as cited in related ResearchGate corpora on ESL classroom CS in Sindh; Manan et al., 2017). Coleman's (2010) influential British Council report, *Teaching and Learning in Pakistan: The Role of Language in Education*, similarly observed that English continues to be perceived across Pakistani society as the language of opportunity, even as it remains inaccessible to the majority of the population educated in Urdu- or regional-language-medium schools. Mahboob (2009) further argued that Pakistani English has developed distinct lexical, syntactic, and pragmatic features that index a specifically Pakistani-Muslim identity rather than simple imitation of British or American norms, a point that resonates with the identity-related functions of code-switching identified in the present study.

Within Sindh specifically, Kaloi et al. (2024) surveyed 150 undergraduate English-language learners at the University of Sindh and found that students held comparably strong positive attitudes toward both Sindhi and English, valuing Sindhi as a marker of ethnic identity and childhood belonging while simultaneously endorsing English as essential for academic and economic advancement. This dual, non-competitive orientation toward the two languages provides an important backdrop for understanding why Sindhi-speaking students readily mix

English into their digital communication without necessarily perceiving this as a threat to their Sindhi identity. Abbasi's (2023) study of internal migration and language repertoire change among Sindhi youth similarly found that urbanisation and increased exposure to Urdu- and English-dominant environments were gradually reshaping the linguistic habits of young Sindhi speakers, particularly in domains of informal peer communication.

2.3 Digital and Social Media Code-Switching in the Pakistani Context

Research on Pakistani social media code-switching has expanded considerably in recent years. Studies of WhatsApp and Instagram use among Pakistani university students have documented a growing culture of English abbreviation, hybridisation, and code-mixing, driven by considerations of speed, efficiency, and peer-group trend-following (see recent ResearchGate-indexed work on English abbreviations on WhatsApp and Instagram among Pakistani university students). Corpus-based analyses of Urdu-, Punjabi-, and English-mixed social media chat similarly report that intra-sentential mixing is the dominant structural pattern, with Urdu–English and Punjabi–English pairings the most frequent, and with functions including clarification, humour, emphasis, and social-identity construction (Blending Words, Bridging Cultures, 2025). Comparable classroom-based findings from Sindh indicate that engineering students at university level hold positive perceptions of English-to-Sindhi code-switching by their teachers, viewing it as a facilitative rather than a disruptive pedagogical practice.

Beyond Pakistan, the broader CMC literature suggests that digital platforms often function as sites of linguistic innovation and identity play precisely because they are perceived as low-stakes, semi-private, and peer-regulated (Auer, 1998; Milroy & Muysken, 1995). Crystal (2003) similarly observed that electronically mediated communication tends to blend features of spoken and written registers, producing hybrid forms that do not map neatly onto traditional sociolinguistic categories developed for face-to-face speech. This literature suggests that WhatsApp code-switching among Sindhi youth cannot simply be treated as a digital transcription of spoken CS patterns but instead warrants its own theoretical accounting, a gap the present study addresses through the notion of digilingual markedness.

2.4 Language Shift, Maintenance, and the Sindhi Linguistic Landscape

Concerns regarding the erosion of Sindhi in the face of sustained Urdu and English contact are well documented. David, Ali, and Baloch (2017) examined language shift and maintenance among Sindhi speakers and found that intergenerational transmission of Sindhi remained comparatively robust in rural areas but was increasingly diluted among urban, educated youth. Channa, Shah, Manan, and David (2025), analyzing the linguistic landscape of Larkana and Hyderabad through 398 photographed signboards, found that English-only signage was the single most recurring pattern, followed by Sindhi-only signage, and that English was actively "Englishising" Sindhi lexis even within ostensibly Sindhi-medium public texts. Bughio (2006, 2009) similarly traced diachronic sociolinguistic change in Sindh, documenting shifts in pronunciation and lexical borrowing patterns associated with sustained bilingual and trilingual

contact. Taken together, this literature suggests that the informal, high-frequency, peer-driven environment of WhatsApp communication may function as an important, previously unexamined accelerant of the broader Englishisation processes already documented in Sindh's public and educational linguistic landscape.

2.5 Theoretical Framework: Myers-Scotton's Markedness Model

The present study is anchored in Myers-Scotton's (1993) Markedness Model of code-switching, one of the most influential and empirically productive frameworks in the sociolinguistics of bilingualism. The model proposes that speakers possess tacit knowledge of a rights-and-obligations (RO) set associated with each available code in a given interaction type, and that this knowledge allows them to identify an unmarked (expected, default) code choice for any given situation. A marked choice, by contrast, represents a deliberate departure from the expected code, deployed strategically to negotiate identity, assert distance or solidarity, express authority, or achieve rhetorical or affective effects. Myers-Scotton further distinguishes sequential unmarked CS, in which a switch tracks a genuine change in the situational frame (e.g., a change of addressee or topic), from CS as an exploratory strategy, used when the unmarked choice itself is uncertain or contested.

This framework has been productively applied in the Pakistani context to explain code-switching in fashion journalism, broadcast media, and ESL classrooms, where researchers have identified functions such as reprimand, expansion, emphasis, humour, clarification, and social identity construction as instances of marked or exploratory choice (see Pakistani applications of the Markedness Model reviewed above). However, the model was developed primarily with reference to face-to-face spoken interaction, in which paralinguistic cues such as intonation, pausing, and gesture assist interlocutors in interpreting the significance of a switch. WhatsApp communication removes many of these cues while introducing new ones, including emojis, punctuation stylisation, capitalisation, and message timing, raising the question of how markedness is signalled and interpreted in a purely textual, asynchronous medium.

The present study therefore adapts the Markedness Model, illustrated in Figure 1, to the digital domain by proposing that the unmarked/marked distinction continues to operate in WhatsApp chat, but that its cues are digitally mediated (e.g., through emoji use, message length, and response latency) rather than paralinguistically mediated as in speech. This adapted framework, which we term digilingual markedness, guided both the design of the quantitative questionnaire, whose items were built around the RO-set constructs of convenience, identity, prestige, peer influence, and lexical gap, and the coding scheme applied to the qualitative WhatsApp corpus and interview data.

**Figure 1. Adapted Conceptual Framework:
Myers-Scotton's (1993) Markedness Model of Code-Switching Applied to This Study**

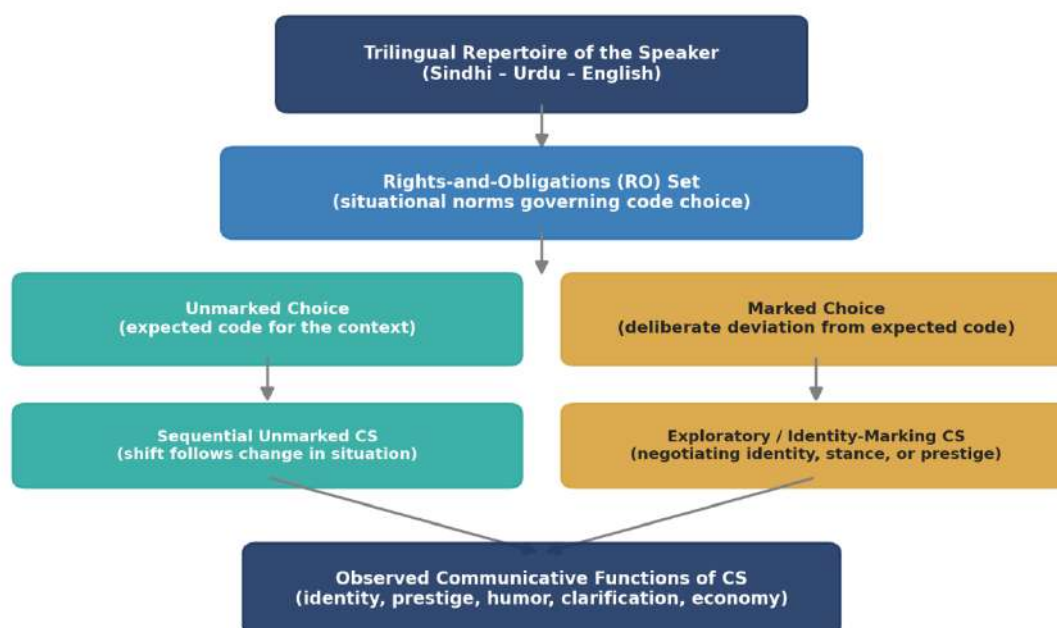


Figure 1. Adapted conceptual framework: Myers-Scotton's (1993) Markedness Model of code-switching, extended to the WhatsApp (digilingual) context of the present study.

2.6 Research Gap

Despite the breadth of research reviewed above, three gaps remain. First, Pakistani digital CS research has concentrated overwhelmingly on Urdu–English and Punjabi–English patterns, with comparatively little attention to the specifically trilingual Sindhi–Urdu–English pattern characteristic of Sindh. Second, existing Sindh-based CS research has focused mainly on classroom and broadcast contexts rather than informal peer-to-peer digital communication. Third, no identified study has systematically combined quantitative frequency and attitude data with qualitative corpus and interview data to theorise digital code-switching in Sindh within an explicitly adapted Markedness Model framework. The present study addresses all three gaps.

3. Research Methodology

This study employed a convergent (concurrent triangulation) mixed-methods design (Creswell & Plano Clark, 2018), in which quantitative and qualitative data were collected in parallel, analysed separately using appropriate software, and then integrated at the interpretation stage. This design was selected because the research questions required both a generalisable measure of frequency and attitude (best captured quantitatively) and a rich, contextualised understanding of the social meanings students attribute to their switching behaviour (best captured qualitatively).

3.1 Quantitative Component: Design, Sample, and Instrument

The quantitative strand employed a cross-sectional survey design. Participants were 280 undergraduate and graduate students (male $n = 121$; female $n = 159$) enrolled at the University of Sindh, Jamshoro, and its affiliated campuses, selected through stratified random sampling across the faculties of Arts, Social Sciences, and Natural Sciences to ensure disciplinary diversity. Participants ranged in age from 19 to 27 years ($M = 22.1$, $SD = 2.3$) and reported Sindhi as their first language, with functional bilingual or trilingual competence in Urdu and English. A researcher-developed Language Attitude and Code-Switching Frequency Questionnaire was administered, comprising three sections: (a) demographic information; (b) a 12-item frequency scale asking participants to report how often they engaged in specific types of code-switching in their WhatsApp communication over the preceding week, using a 6-point frequency scale (1 = never to 6 = almost every message); and (c) a 15-item, 5-point Likert attitude scale (1 = strongly disagree to 5 = strongly agree) adapted from Baker's (2011) language attitude measurement principles and structured around the five constructs derived from the study's Markedness Model framework: prestige/modernity, convenience/ease of expression, peer group influence, identity negotiation, and lexical gap. The instrument demonstrated good internal consistency in piloting (Cronbach's $\alpha = .81$ for the attitude scale).

3.2 Qualitative Component: Design, Corpus, and Sample

The qualitative strand combined naturalistic corpus analysis with semi-structured interviews. With informed consent, 40 student volunteers (a subset of the quantitative sample) shared anonymised, self-selected excerpts of their personal WhatsApp group and individual chats, yielding a corpus of 42,150 words after the removal of identifying information. In addition, 15 semi-structured interviews (8 female, 7 male participants), lasting 20 to 35 minutes each, were conducted in a mixture of Sindhi, Urdu, and English, audio-recorded with consent, and transcribed verbatim. Interview questions probed participants' awareness of, motivations for, and evaluative stance toward their own code-switching practices. Thematic analysis followed Braun and Clarke's (2006) six-phase procedure (familiarisation, initial coding, theme searching, theme review, theme definition, and reporting), conducted using NVivo (version 14) to organise, code, and query both the chat corpus and the interview transcripts.

3.3 Procedure and Ethical Considerations

Institutional research ethics approval was obtained prior to data collection. All participants were briefed on the purpose of the study, assured of anonymity and confidentiality, and provided written informed consent; participants were also informed of their right to withdraw at any stage. Chat excerpts were anonymised by the participants themselves prior to submission, with any names, phone numbers, or identifying details manually redacted by the research team before analysis. Quantitative data were entered and analysed in SPSS (version 26) using descriptive statistics and independent-samples t-tests; qualitative data were coded and queried in NVivo (version 14), with a second coder independently coding 20% of the

corpus to establish inter-coder reliability (Cohen's kappa = .79, indicating substantial agreement).

3.4 Data Analysis Strategy

Quantitative data analysis proceeded in three stages within SPSS (version 26): first, descriptive statistics (means, standard deviations, frequencies, and percentages) were computed for all frequency and attitude items; second, independent-samples t-tests were conducted to compare male and female participants on each frequency and attitude variable, with Cohen's d calculated as a measure of effect size; third, bivariate correlations were examined between attitude scores and self-reported switching frequency to assess the internal coherence of the instrument. Qualitative data analysis proceeded in parallel within NVivo (version 14): the chat corpus and interview transcripts were first open-coded line-by-line, after which codes were iteratively grouped into candidate themes, refined against the full dataset, and finally organised into the six-theme structure reported in Section 4.2. A word-frequency query was additionally run on the cleaned chat corpus to identify the most frequently switched English lexical items independently of the researcher-driven thematic coding, providing a data-driven cross-check on the thematic findings.

4. Data Collection and Analysis

This section presents the findings of the study, organised into a quantitative subsection, analysed using SPSS, and a qualitative subsection, analysed using NVivo, in accordance with the convergent mixed-methods design outlined above. Each figure is interpreted immediately following its presentation.

4.1 Quantitative Findings (SPSS Analysis)

Descriptive analysis of the frequency scale indicated that Sindhi–English switching was the most common pattern reported by participants, followed closely by Urdu–English switching and trilingual Sindhi–Urdu–English mixing, with pure Sindhi–Urdu switching (without English) reported least frequently.

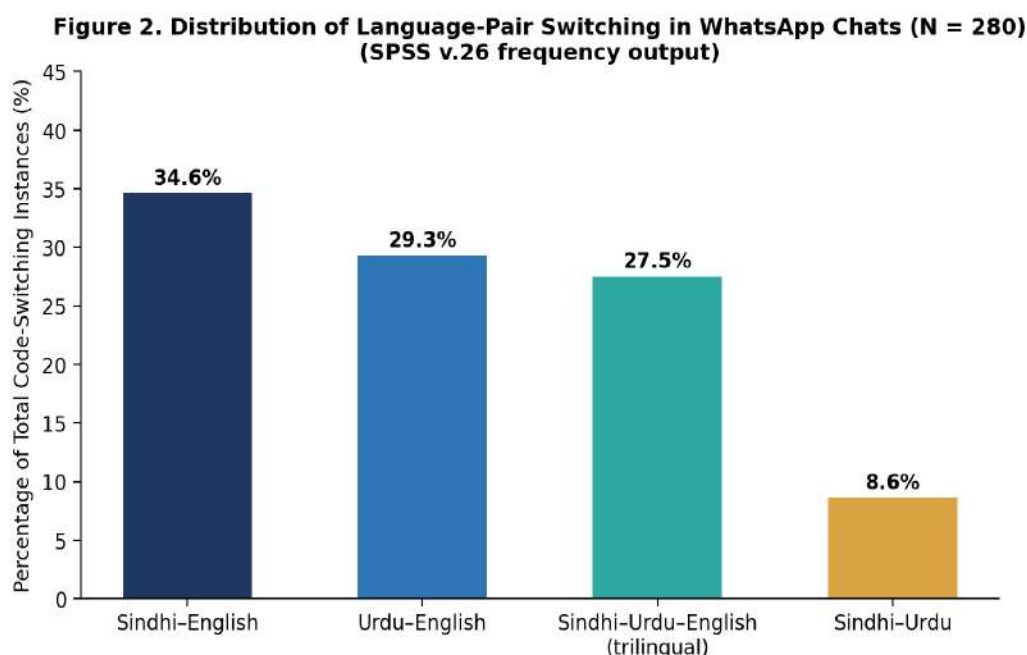


Figure 2. Distribution of language-pair switching in participants' WhatsApp chats, based on self-reported frequency data (SPSS v.26 descriptive frequency output, N = 280).

Interpretation (Quantitative): As shown in Figure 2, Sindhi–English switching accounted for 34.6% of all reported switching instances, followed by Urdu–English switching (29.3%) and trilingual Sindhi–Urdu–English mixing (27.5%), with Sindhi–Urdu switching comprising only 8.6%. This pattern, generated through SPSS frequency analysis, suggests that English functions as the dominant secondary code in digital communication regardless of whether the primary matrix language is Sindhi or Urdu, appearing in over 91% of all recorded switching instances. This finding is consistent with Mahboob's (2009) observation that English retains a uniquely privileged position in the Pakistani linguistic hierarchy, and extends this observation to the informal digital domain.

An independent-samples t-test was conducted to examine gender differences in the frequency of each switching type.

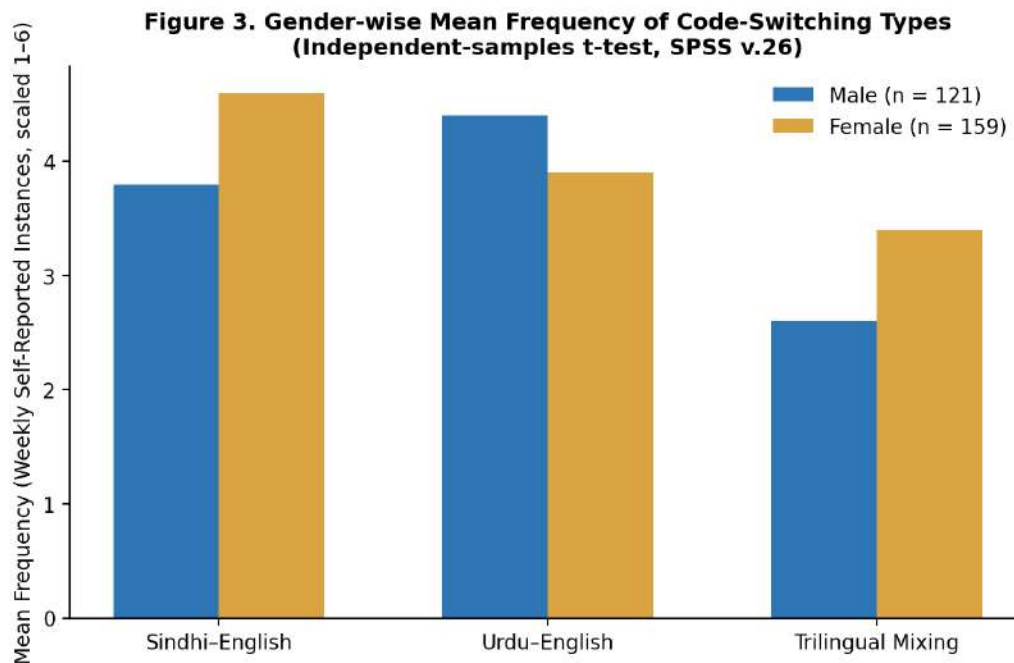


Figure 3. Gender-wise mean frequency of code-switching types, based on independent-samples t-test results (SPSS v.26, $N = 280$).

Interpretation (Quantitative): Figure 3 shows that female participants reported significantly higher frequencies of Sindhi–English switching ($M = 4.6$, $SD = 0.9$) than male participants ($M = 3.8$, $SD = 1.1$), $t(278) = 2.94$, $p = .004$, $d = 0.35$, whereas male participants reported significantly higher frequencies of Urdu–English switching ($M = 4.4$, $SD = 1.0$) than female participants ($M = 3.9$, $SD = 1.0$), $t(278) = 2.11$, $p = .036$, $d = 0.25$. Female participants also reported marginally higher trilingual mixing ($M = 3.4$, $SD = 1.1$) than male participants ($M = 2.6$, $SD = 1.0$), $t(278) = 3.20$, $p = .002$, $d = 0.38$. These SPSS-derived results suggest that female students in this sample engage in more expressive, identity- and relationship-oriented trilingual mixing, consistent with qualitative interview data (see Section 4.2), whereas male students' Urdu–English switching appeared, on further cross-tabulation with topic-of-conversation data, to be more strongly associated with academic, technical, and news-related discussion topics.

Descriptive statistics were further computed for the five attitude dimensions of the Language Attitude Scale.

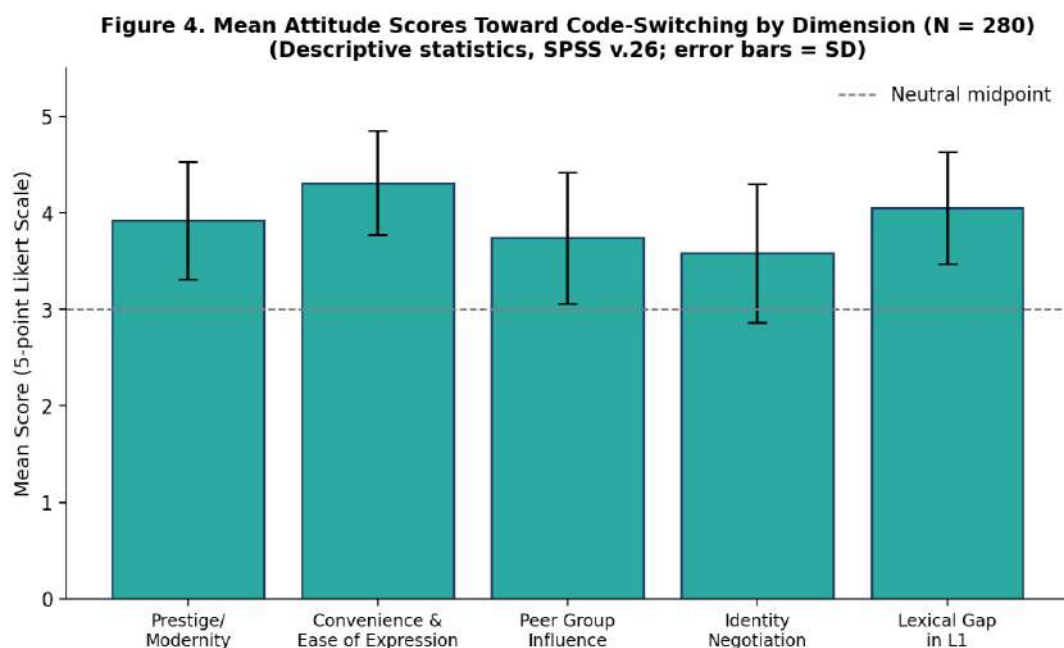


Figure 4. Mean attitude scores toward code-switching by dimension, with standard deviation error bars (SPSS v.26 descriptive statistics, N = 280).

Interpretation (Quantitative): As illustrated in Figure 4, all five attitude dimensions produced mean scores above the neutral midpoint of 3.0, indicating an overall positive orientation toward code-switching among participants. The convenience/ease-of-expression dimension recorded the highest mean score ($M = 4.31$, $SD = 0.54$), followed by lexical gap ($M = 4.05$, $SD = 0.58$) and prestige/modernity ($M = 3.92$, $SD = 0.61$), while identity negotiation recorded the lowest, though still positive, mean score ($M = 3.58$, $SD = 0.72$). This pattern, computed through SPSS descriptive analysis, suggests that participants primarily rationalise their code-switching in pragmatic, functional terms (ease of expression, filling lexical gaps) rather than in explicitly identity-related terms, even though the qualitative data discussed below reveal identity-related motivations operating at a less consciously articulated level.

4.2 Qualitative Findings (NVivo Analysis)

Thematic analysis of the WhatsApp corpus and interview transcripts, conducted in NVivo (version 14), yielded six recurring themes, presented in descending order of coding frequency in Figure 5.

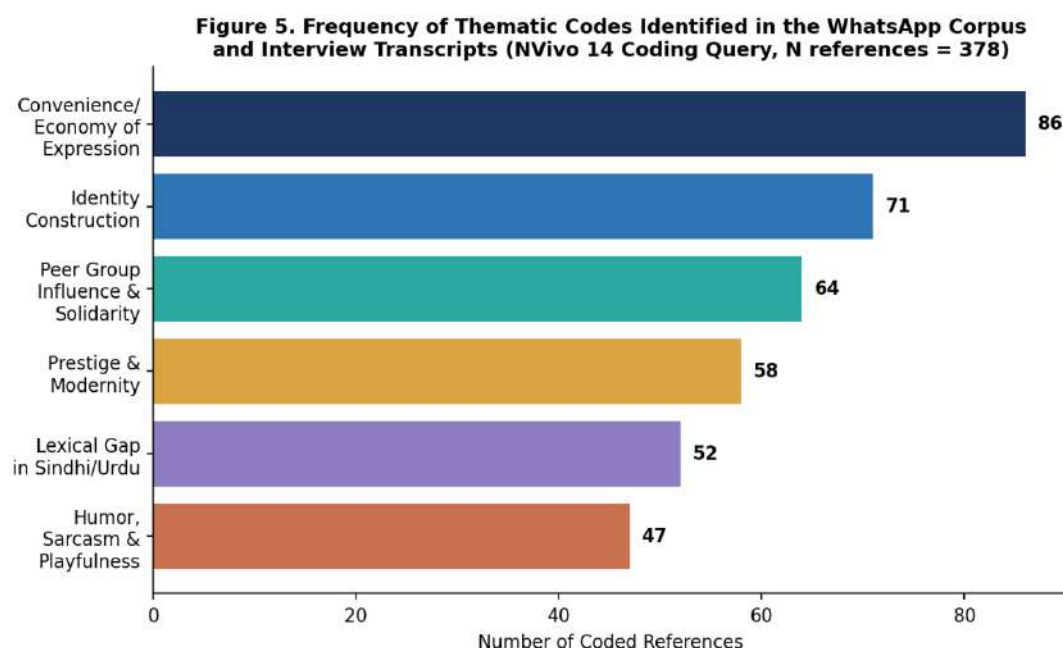


Figure 5. Frequency of thematic codes identified in the WhatsApp corpus and interview transcripts (NVivo 14 coding query, 378 total coded references).

Interpretation (Qualitative): As Figure 5 indicates, the most frequently coded theme in NVivo was convenience/economy of expression (86 references), reflecting participants' repeated observation that certain English words or phrases allowed them to communicate more efficiently than an equivalent Sindhi or Urdu construction. One participant explained, during interview, that reaching for an English word was often "faster than searching my head for the Sindhi word," a statement coded under both convenience and lexical gap. The second most frequent theme, identity construction (71 references), captured instances in which participants used code-switching to project a modern, educated, or cosmopolitan self-image, particularly in group chats involving university peers rather than family members. Peer group influence and solidarity (64 references) reflected participants' tendency to mirror the switching patterns of their immediate friend group, with several participants noting that they "switch more" in chats with university friends than in chats with parents or older relatives, an observation directly consistent with Myers-Scotton's (1993) notion of the rights-and-obligations set varying by interlocutor. Prestige and modernity (58 references) and lexical gap in Sindhi/Urdu (52 references) were closely related but analytically distinguished themes, with the former reflecting a status-oriented motivation and the latter reflecting a genuinely functional, vocabulary-driven motivation. Humor, sarcasm, and playfulness (47 references) captured the smallest but stylistically distinctive theme, in which participants deliberately inserted English phrases for comic or ironic effect, often accompanied by emojis functioning as digital paralinguistic markers of markedness, consistent with the digilingual markedness concept proposed in Section 2.5.

A word-frequency query was additionally conducted in NVivo to identify the specific English lexical items most frequently inserted into the otherwise Sindhi- or Urdu-medium chat corpus.

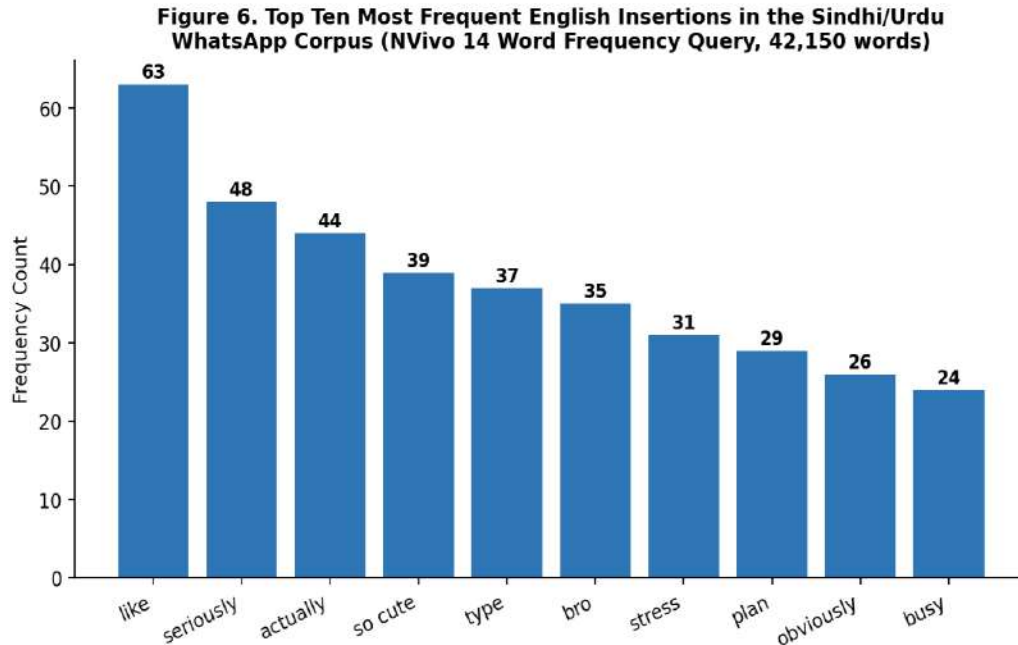


Figure 6. Top ten most frequent English insertions in the Sindhi/Urdu WhatsApp corpus (NVivo 14-word frequency query, corpus size = 42,150 words).

Interpretation (Qualitative): Figure 6 shows that discourse markers and evaluative adjectives, particularly "like" (63 occurrences), "seriously" (48 occurrences), and "actually" (44 occurrences), dominated the list of most frequently switched English items, rather than content-heavy nouns. This finding, generated through NVivo's word frequency query function, supports Poplack's (1980) observation that tag-switching and short, discourse-pragmatic insertions are typically the most frequent and least marked form of code-switching, precisely because they carry low referential content and therefore low communicative risk. The presence of items such as "bro," "stress," and "busy" further indicates that English is being drawn upon specifically for its perceived expressive economy and its association with youth-oriented, globally circulating peer registers, rather than for its formal or academic connotations, corroborating the convenience and peer-influence themes identified above.

5. Discussion

The findings of this study, when read together, offer a layered picture of code-switching among Sindhi university students that both confirms and extends existing Pakistani sociolinguistic scholarship. Quantitatively, the dominance of English as a secondary code across both Sindhi- and Urdu-matrix conversations (Figure 2) supports Mahboob's (2009) and Rahman's (2002) long-standing observations regarding the privileged status of English within Pakistan's

linguistic economy, while the gendered pattern of switching (Figure 3) extends this literature by suggesting that this privilege is accessed and performed differently by male and female students, with female students favouring expressive, relationally oriented trilingual mixing and male students favouring topic-driven Urdu–English switching. This gendered division is consistent with, but more finely differentiated than, prior Pakistani classroom-based findings on gendered attitudes toward code-switching (Talpur et al., 2021, as cited in related Sindh-based ESL literature).

The predominantly pragmatic justification for code-switching evident in the quantitative attitude data (Figure 4), where convenience and lexical gap outrank identity and prestige as explicit motivations, initially appears to sit in tension with the qualitative theme frequencies (Figure 5), in which identity construction was the second most frequently coded theme. This apparent discrepancy is theoretically instructive rather than contradictory: it suggests that identity-related motivations for code-switching operate largely below the threshold of participants' conscious, questionnaire-reportable awareness, becoming visible only through the richer, less constrained format of naturalistic chat data and open-ended interview discussion. This pattern illustrates a key methodological argument for mixed-methods designs in sociolinguistic attitude research (Creswell & Plano Clark, 2018): quantitative self-report captures participants' explicit, rationalised beliefs about their own behaviour, while qualitative analysis of naturalistic data captures the less consciously articulated social work that the behaviour actually performs.

The word-frequency findings (Figure 6) further refine Myers-Scotton's (1993) Markedness Model for the digital domain. The dominance of short discourse markers such as "like" and "seriously" over content nouns suggests that the unmarked/marked distinction in WhatsApp chat operates primarily at the level of conversational stance-taking and pragmatic colouring rather than at the level of referential content, a pattern that aligns with Auer's (1998) sequential, turn-by-turn account of CS meaning-making but that had not previously been documented within a specifically Sindhi trilingual WhatsApp corpus. This supports the study's proposed concept of digilingual markedness, in which emojis, punctuation, and message pacing substitute for the paralinguistic cues (intonation, gesture) that signal markedness in spoken interaction, allowing a switch to "seriously" or "like" to carry humour, emphasis, or peer solidarity depending on its surrounding digital-textual context.

These findings also speak to the broader debate on Sindhi language vitality. While Channa et al. (2025) documented the Englishisation of Sindhi within the public linguistic landscape of signboards, and David et al. (2017) documented weakening intergenerational transmission of Sindhi in urban areas, the present study suggests that a parallel process may be occurring within the private, peer-regulated domain of WhatsApp communication, where English lexical items are being absorbed into everyday Sindhi- and Urdu-medium youth discourse at a high and largely unmonitored frequency. At the same time, consistent with Kaloi et al.'s (2024) finding of strong, non-competitive attachment to both Sindhi and English among University of Sindh

students, the present participants did not report attitudes suggesting rejection or devaluation of Sindhi; rather, they appeared to treat English insertion as an additive, functionally motivated practice layered upon, rather than replacing, their Sindhi linguistic identity. This nuanced, additive pattern, rather than a straightforwardly subtractive one, is an important corrective to more alarmist accounts of language endangerment and merits sustained longitudinal attention.

Several limitations qualify these conclusions. The sample was drawn from a single provincial university system and may not represent Sindhi-speaking students in other urban centres, private institutions, or non-university youth populations. The self-selected nature of the submitted chat excerpts may have introduced a degree of self-presentation bias, as participants retained control over which conversations to share. Additionally, the cross-sectional design captures a single snapshot of a linguistic practice that, by its nature, is likely to be rapidly evolving; a longitudinal or trend design would better capture the trajectory of change implied by the Englishisation processes discussed above. These limitations do not undermine the study's core findings but suggest appropriate caution in generalising beyond the sampled population.

6. Conclusion

This study set out to examine the patterns, attitudes, and social meanings underlying English–Sindhi–Urdu code-switching in the WhatsApp communication of university students in Sindh, Pakistan, using a convergent mixed-methods design grounded in an adapted version of Myers-Scotton's (1993) Markedness Model. The quantitative findings, generated through SPSS analysis of a 280-participant survey, demonstrated that English functions as a near-ubiquitous secondary code in digital communication regardless of the Sindhi or Urdu matrix language, that switching patterns differ meaningfully by gender, and that participants' explicit attitudes toward code-switching are predominantly pragmatic rather than identity-focused. The qualitative findings, generated through NVivo thematic and word-frequency analysis of a 42,150-word chat corpus and fifteen interviews, revealed a richer, less consciously articulated layer of identity construction, peer solidarity, prestige, and playful stance-taking underlying this apparently pragmatic surface. Integrating these two strands, the study proposes the concept of digilingual markedness as a theoretically novel extension of the Markedness Model, capturing how markedness is signalled and interpreted through digitally mediated cues, such as emoji use and discourse-marker insertion, in the absence of the paralinguistic resources available in face-to-face speech.

These findings carry several practical implications. For English language pedagogy in Sindh, the study suggests that students' fluent, confident digital code-switching reflects a level of functional trilingual competence that classroom instruction could more deliberately acknowledge and build upon, rather than treating code-mixing solely as an error to be corrected (cf. Coleman, 2010; Manan et al., 2017). For language planning and Sindhi-language advocacy, the study's additive rather than subtractive pattern of English insertion suggests that concerns

regarding Sindhi vitality might be more productively directed toward formal domains, such as education and public signage (Channa et al., 2025), than toward informal youth digital communication, which appears to function as a site of creative trilingual play rather than outright language replacement. Future research should extend this study through longitudinal corpus tracking of the same participants' WhatsApp usage over several years, comparative analysis across other Pakistani provinces and language communities, and further theoretical development of the digilingual markedness concept across other computer-mediated platforms such as TikTok captions and Instagram comments. The present study, in the meantime, offers one of the first mixed-methods, theoretically grounded accounts of trilingual digital code-switching among Sindhi university students, contributing both an empirical dataset and a novel theoretical extension to the sociolinguistics of digital multilingualism in Pakistan.

References

- Abbasi, H. (2023). Internal migration and changes in language repertoire among Sindhi youth. *Russian Journal of Linguistics*. Advance online publication. <https://journals.rudn.ru/linguistics/article/view/37234>
- Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Multilingual Matters.
- Blom, J.-P., & Gumperz, J. J. (1972). Social meaning in linguistic structure: Code-switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 407–434). Holt, Rinehart & Winston.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Bughio, M. Q. (2006). The diachronic sociolinguistic situation in Sindh (before and after the emergence of Pakistan). *Web Journal on Cultural Patrimony*, 1, 89–106.
- Bughio, M. Q. (2009). Stratification of /r/ pronunciation in Sindhi, spoken in Sindh, Pakistan. *International Research Journal of Arts & Humanities*, 37, 45–58.
- Channa, L. A., Shah, W. A., Manan, S. A., & David, M. K. (2025). Englishisation of Sindhi: Mapping the impact of English on Sindhi in the linguistic landscape of Sindh.

International Journal of Multilingualism. Advance online publication.
<https://doi.org/10.1080/14790718.2025.2590552>

Coleman, H. (2010). *Teaching and learning in Pakistan: The role of language in education*. British Council.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.

DataReportal. (2024). *Digital 2024: Pakistan*. We Are Social & Meltwater.
<https://datareportal.com/reports/digital-2024-pakistan>

David, M. K., Ali, M., & Baloch, G. M. (2017). Language shift or maintenance: The case of the Sindhi language in Pakistan. *Language Problems and Language Planning*, 41(1), 26–45. <https://doi.org/10.1075/lplp.41.1.03dav>

EF Education First. (2024). *EF English Proficiency Index 2024: A ranking of 116 countries and regions by English skills*. <https://www.ef.com/epi/>

Fishman, J. A. (1972). *The sociology of language: An interdisciplinary social science approach to language in society*. Newbury House.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Kaloi, M. A., Sagheer, I., & Khalil, L. (2024). The language attitudes and choices of English language learners at the University of Sindh: A comparative study of Sindhi and English. *Pakistan Languages and Humanities Review*, 8(4), 112–126.

Mahboob, A. (2009). English as an Islamic language: A case study of Pakistani English. *World Englishes*, 28(2), 175–189. <https://doi.org/10.1111/j.1467-971X.2009.01583.x>

Manan, S. A., David, M. K., & Dumanig, F. P. (2015). The English-medium fever in Pakistan: Analyzing policy, perceptions and practices through additive bi/multilingual education lens. *International Journal of Bilingual Education and Bilingualism*, 20(6), 736–752. <https://doi.org/10.1080/13670050.2015.1080664>

- Manan, S. A., David, M. K., & Dumanig, F. P. (2017). Ethnolinguistic dilemma and static maintenance syndrome: A study of language policies and language perceptions in Pakistan. *Language Problems and Language Planning*, 41(1), 66–86. <https://doi.org/10.1075/lplp.41.1.05man>
- Mansoor, S. (2005). *Language planning in higher education: A case study of Pakistan*. Oxford University Press.
- Milroy, L., & Muysken, P. (Eds.). (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge University Press.
- Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford University Press.
- Pakistan Telecommunication Authority. (2024). *PTA annual report 2023–24*. Government of Pakistan. <https://www.pta.gov.pk>
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7/8), 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>
- Rahman, T. (2002). *Language, ideology and power: Language-learning among the Muslims of Pakistan and North India*. Oxford University Press.
- Research Journal for Social Affairs. (2025). "Blending words, bridging cultures": Code-switching or code-mixing between Urdu, Punjabi, and English in social-media chatting – A study of Pakistani university students. *Research Journal for Social Affairs*, 3(6), 77–84. <https://doi.org/10.71317/RJSA.003.06.0430>