Running head: CHINESE ACE-IQ

Examining Patterns of Adversity in Chinese Young Adults Using the Adverse Childhood Experiences – International Questionnaire (ACE-IQ)

Grace W.K. Ho, PhD, RN School of Nursing, The Hong Kong Polytechnic University

Athena C.Y. Chan, MS School of Nursing, The Hong Kong Polytechnic University

Wai-Tong Chien, PhD, M.Phil, BN The Nethersole School of Nursing, The Chinese University of Hong Kong

> Daniel T. Bressington, PhD, RN School of Nursing, The Hong Kong Polytechnic University

> > Thanos Karatzias, PhD Edinburgh Napier University; NHS Lothian

Funding

This work was supported by The Hong Kong Polytechnic University Internal Start-Up Fund.

Conflict of Interest None to declare.

Corresponding Author: Grace W.K. Ho, PhD, RN PQ426, School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, HKSAREmail:grace.wk.ho@polyu.edu.hk | Tel: +852 3400 3918 Running head: Chinese ACE-IQ

Abstract

Ample evidence supports significant and enduring associations between adverse childhood experiences (ACEs) and negative outcomes later in life. Subsets of ACEs (e.g. childhood maltreatment and household dysfunction) have been examined in Chinese populations, but no known study has comprehensively examined the full constellation of different types of ACEs or patterns of ACE exposure in Chinese samples. As a direct response to the call to establish a global ACEs surveillance framework, this study provides the first translation and validation of the World Health Organization ACE – International Questionnaire (ACE-IQ). Further, patterns of ACE exposure were identified through latent class analysis. The 29-item ACE-IQ was translated and back-translated from English to traditional Chinese to measure exposure to 13 categories of ACEs. The Chinese ACE-IQ demonstrated good content validity; the ACE-IQ domain subscales also showed satisfactory test-retest reliability and semantic equivalence. In a sample of 433 Chinese young adults, three patterns of ACE exposure were uncovered: Low ACEs (65.82%), Household Violence (24.94%), and Multiple ACEs (9.24%). Concurrent exposure to physical abuse, domestic violence, and emotional abuse (i.e. Household *Violence*) was a novel pattern found in this study sample, and suggests there may be traditional Chinese norms that potentiate risks for violent household environments in the absence of other household risk factors. Findings underscore the importance of examining ACE exposure within local contexts, as children's adverse experiences may be idiosyncratic to geographic, social, and cultural norms.

Keywords: Adverse Childhood Experiences, Adverse Childhood Experiences – International Questionnaire, Psychometric Evaluation, Latent Class Analysis

1

Background

Adverse Childhood Experiences (ACEs) refer to a broad set of negative childhood experiences that include abuse (emotional, physical, or sexual), neglect (emotional or physical), serious household dysfunction (e.g. witnessing domestic violence, household member drug use, and parental separation and incarceration), and peer, community, and collective violence (World Health Organization [WHO], 2016). These early life adversities are pervasive across social and cultural settings (R. C. Kessler et al., 2010), and have been linked with negative health and sociobehavioral outcomes throughout the life course (Shonkoff et al., 2012). The dose-dependent relationships between the number of ACEs and increased risks for negative outcomes are also well-documented (Felitti et al., 1998). However, comparatively little research has comprehensively examined exposure to the full constellation (i.e. 13 categories) of ACEs, especially in ethnically Chinese populations. Further, only a few studies have examined the profiles of ACE exposure, none of which was conducted outside of a Western context. To propel this area of research, this present study was the first to translate and evaluate the psychometric properties and response patterns of a global ACE measure, the ACE-International Questionnaire (ACE-IQ) (World Health Organization [WHO], 2016), in a Chinese sample.

ACEs in Chinese populations

Ample evidence shows that childhood adversities are prevalent among Chinese people, but most existing studies are confined to understanding specific subsets of adversity (i.e. childhood maltreatment, household dysfunction, or violence outside the home). For example, in a representative household survey of over 1,000 children between ages 12-17 in Hong Kong, 36% reported they have experienced physical assault by their mother in the past, psychological aggression and neglect by mother were also reported in 61% and 27% of these children, respectively (Chan, 2012). Similarly, a recent meta-analysis of 68 child maltreatment studies conducted in China estimated that 27% of children were victimized by physical abuse, 20% reported emotional abuse, 9% reported sexual abuse, and 26% reported neglect (Fang et al., 2015). Other non-maltreatment adversities are also pervasive. In a prevalence survey of over 5,000 adults from two metropolitan cities in China, nearly 1 in 3 adults experienced at least one type of serious household dysfunction during childhood (Lee et al., 2011). A representative sample of 18,340 Chinese adolescents aged 15-17 also showed that over 70% reported they have been victims of crime or violence, with 14% being poly-victims (Chan, 2013).

To our knowledge, there are only eight studies published in the English literature within the past two decades that measured ACEs using Chinese samples. Although findings from these studies showed that ACEs are indeed prevalent, the rates of exposure to at least one ACE varied widely from 31% (Wei, 2013) to 94% (Li, Cao, Cao, & Liu, 2015). One possible reason for the large discrepancies in reported rates is the diverse and convenient participant samples, but another potentially more compelling reason is the differences in the inclusion and measurement of different forms of ACEs across studies. For example, across the eight studies, four omitted ACEs related to peer, community, and collective violence (Liu, Yang, Shi, Liu, & Wang, 2016; Wong, Fong, & Chan, 2015; Xiao, Dong, Yao, Li, & Ye, 2008; Xue, Lin, Sun, & Cao, 2017); two also excluded physical and emotional neglect (Ding, Lin, Zhou, Yan, & He, 2014; Wei, 2013), and one also excluded sexual abuse (Xue, Gao, & Cao, 2016). One study further reduced ACEs into four broad categories and reported on the overall prevalence of family, community, school, and personal adversity (Li et al., 2015).

How ACEs were measured also varied widely across Chinese ACE studies. For example, three studies used questions adapted from life events or trauma scales (Li et al., 2015; Wei, 2013;

Xiao et al., 2008), one study measured ACEs using dichotomous responses (i.e. yes or no) (Xue et al., 2016), and the remaining studies did not clearly report the response options or scoring algorithms to derive at ACE exposure. Importantly, none of these studies has reported psychometric properties on the translated or adapted measures, and there is no clear consensus on how ACE exposure was established. To address these gaps, the present study translated and validated an international ACE questionnaire, the ACE-IQ, to generate a standard Chinese measure to comprehensively assess 13 categories of ACEs.

Patterns of ACE exposure

ACEs have a lasting impact on children, potentially setting them on a lifelong trajectory for impaired physical, mental, and behavioral health (Shonkoff et al., 2012). The dose-response relationships between ACE exposure and negative health outcomes in adulthood are also clear. Since the seminal study by Felitti and colleagues (1998), the graded impact of ACEs on risks for chronic diseases (e.g. ischemic heart disease, cancer, and chronic lung and liver diseases), poorer mental health (e.g. psychopathology and suicide attempts), and engaging in health risk behaviors (e.g. smoking, physical inactivity, and \geq 50 sexual partners) have been replicated across diverse populations. Similarly, the limited research conducted in Chinese populations has shown that greater ACE exposure is associated with increased risks for poor health outcomes in adulthood, including alcohol misuse (Liu et al., 2016; Xiao et al., 2008), somatic symptoms (Wong et al., 2015), and drug-induced psychosis (Ding et al., 2014).

However, a unidimensional approach to assessing cumulative ACE exposure (i.e. the ACE score) precludes understanding possible heterogeneity or typologies of victims with different patterns of ACEs who may, in turn, have different risk profiles for different types of negative outcomes (Cavanaugh, Petras, & Martins, 2015). Indeed, studies that examined ACEs

using a pattern-based approach demonstrated that different ACE profiles confer different risks for psychological disturbance, substance use, and criminality in adulthood (Cavanaugh et al., 2015; L. E. Roos et al., 2016; Shin, McDonald, & Conley, 2018). Understanding these risk profiles have important utility in tailoring interventions to prevent and mitigate the deleterious outcomes associated with different typologies of exposure. Additionally, patterns of ACE must be examined within the context of culture, as definitions, occurrence, and co-occurrence of ACEs vary widely across geographic boundaries and socioeconomic norms (R. C. Kessler et al., 2010; Stoltenborgh, Bakermans-Kranenburg, van Ijzendoorn, & Alink, 2013). To this end, this study examined ACEs in a sample of young Chinese adults using a pattern-based approach, which was the first to identify typologies of ACE exposure outside a Western context.

The current study

This study was conducted as a direct response to the call by the World Health Organization to build a global ACEs surveillance framework to comprehensively examine ACE exposure across populations using the ACE-IQ (Anda, Butchart, Felitti, & Brown, 2010). In a sample of young adults in Hong Kong, the main objectives of this study were to: (1) provide translation and content validation of the Chinese ACE-IQ; (2) evaluate the psychometric properties of the Chinese ACE-IQ (i.e. internal consistency, test-retest reliability, and semantic equivalence with the English version); and (3) investigate patterns of ACE exposure based on 13 categories of ACEs.

Methods

This study was conducted in two phases. In the first phase, the ACE-IQ was translated and back-translated from English to traditional Chinese; an expert panel provided content validation of the translated measure based on its relevance and appropriateness in the Chinese culture. In the second phase, the Chinese ACE-IQ was administered to a sample of young adults in Hong Kong. Test-retest reliability and semantic equivalence were assessed in a subgroup of participants. Lastly, Patterns of ACE were examined using latent class analysis (LCA). Ethical approval was obtained from the first author's university ethics review board.

Study Measure

The ACE-IQ (World Health Organization [WHO], 2016) is a 29-item measure that assesses exposure to 3 domains of childhood adversities – "childhood maltreatment", "family/ household dysfunction", and "violence outside the home" (see Table 1). Collectively, these domains represent 13 categories of ACEs. Childhood maltreatment includes emotional neglect (2 items, P1-P2); physical neglect (3 items, P3-P5); emotional abuse (2 items, A1-A2); physical abuse (2 items, A3-A4); and sexual abuse (4 items; A5-A8). Family/ household dysfunction includes living with substance abuser (1 item, F1); living with household member who was mentally ill or suicidal (1 item, F2); living with household member who was imprisoned (1 item, F3); parental death, separation, or divorce (2 items, F4-F5); and domestic violence (3 items, F6-F8). Violence outside the home includes bullying (1 item, V1); witnessed community violence (3 items, V4-V6); and exposure to war/ collective violence (4 items, V6-V10).

Respondents are asked to respond to the questions based on their experiences during the first 18 years of their lives. Response options for each question may be dichotomous (i.e. Yes/ No; Items F1-F5), based on a 5-point Likert scale ranging from "Never" to "Always" (Items P1-P2), or based on a 4-point Likert scale ranging from "Never" to "Many times" (all remaining items). The original scale developers have proposed two scoring algorithms. First, the binary scoring method uses the lowest threshold for identifying ACEs, where any experience of the adversity denotes exposure (e.g. being screamed or sweared at *once* constitutes an affirmative

response for emotional abuse). A second scoring method, the frequency method, accounts for the level of exposure, which differ by ACE type (e.g. exposure to contact sexual abuse only requires being touched in a sexual way *once*, whereas exposure to emotional abuse requires being screamed or sweared at *many times*) (World Health Organization [WHO], 2016).

In this study, the frequency scoring method was used to generate a conservative estimate of ACE exposure that more closely approximates international norms. Additionally, scoring of 2 ACE categories (i.e. emotional neglect and physical abuse) were adjusted based on proposed changes from prior studies that tested the ACE-IQ in two culturally similar Asian countries (i.e. Vietnam and South Korea) (Kim, 2017; Q. A. Tran, M. P. Dunne, T. V. Vo, & N. H. Luu, 2015). Specifically, an affirmative response to items measuring emotional neglect requires report of "Never"; an affirmative response to items measuring physical abuse requires report of "A Few Times" or "Many Times."

The ACE-IQ measures exposure to 13 categories of ACEs, which can be combined to describe the total number of different ACEs an individual was exposed to. Exposure to an ACE category requires an affirmative response to at least one of the items under that category (e.g., exposure to emotional neglect requires a response of "Never" for at least one of two items on emotional neglect). Thus, initial scoring for each ACE category determines if the participant is "exposed" or "not exposed" to that ACE. Then, the total number of ACE categories that the participant was "exposed" to are summed to create an ACE score ranging from 0-13. This study reports findings on participants' exposure to individual ACE categories and their overall ACE exposure (i.e. the ACE score). Overall, the ACE-IQ demonstrated good internal consistency in the present study, with a Cronbach's alpha of 0.83. The Cronbach's alpha for the "childhood

maltreatment", "family/ household dysfunction", and "violence outside the home" domain subscales were 0.74, 0.62, and 0.60, respectively.

Phase 1: Translation and Content Validation

The ACE-IQ was translated from English to traditional Chinese by a bilingual technical writer, and back-translated by a bilingual study team member. Three other study team members who are experienced in family violence and mental health research independently reviewed the initial translations and provided comments for revision. Comments were compiled and shared among all team members; two additional iterations of the translations were generated before a consensus was reached for the initial draft.

Items from the initial draft were assessed for content validity based on their relevance to childhood adversities and their appropriateness in Chinese culture and society (Polit & Beck, 2006; Polit, Beck, & Owen, 2007). An expert panel of 2 mental health nurse researchers, 2 clinical psychologists, and 2 social workers who regularly engage with vulnerable families and who were not part of the study team were invited to rate the relevance and appropriateness of each translated question. Each questionnaire item was rated on a 4-point Likert scale, ranging from "4 - Highly relevant / appropriate," "3 - Quite relevant / appropriate," "2 - Somewhat relevant / appropriate," to "1 - Not relevant / appropriate."

Content validity indices were calculated for relevance and appropriateness both by item and for the overall questionnaire. Item-level analysis was conducted using item content validity index (I-CVI), i.e. the proportion of experts who gave the item a score of 3 or 4 (Polit & Beck, 2006). A modified kappa statistic (k*) was computed to correct the chance agreement among experts that may artificially inflate I-CVI ratings (Polit et al., 2007). Content validity index for the overall scale (S-CVI) was computed using two methods: (1) universal agreement (S-CVI_{UA}), i.e. the proportion of items that received a score of 3 or 4 by all experts, and (2) an average I-CVI of all scale items (S-CVI_{Ave}) (Polit & Beck, 2006). I-CVI, S-CVI_{UA}, and S-CVI_{Ave} are considered good when coefficient exceeds 0.78, 0.80, and 0.90, respectively under a panel of six to eight raters (Lynn, 1986; Polit & Beck, 2006); $k^* > 0.74$ is considered excellent (Polit et al., 2007).

Overall, the translated items for the ACE-IQ received excellent ratings on relevance and appropriateness, with I-CVIs ranging between 0.83 to 1.0, and k* between 0.82 to 1.0. The scale-level content validity was also high, with S-CVI_{UA} and S-CVI_{Ave} for relevance and appropriateness ranging from 0.86 to 0.99. Of note, one expert commented that the questions related to collective violence (e.g. "Was a family member or friend killed or beaten up by soldiers, police, militia or gangs?") may not be relevant to the childhood experiences of young adults born and raised in Hong Kong. However, these items were retained for purposes of future comparisons with findings from other countries using the same measure. Since soldiers and militia are not applicable in Hong Kong, these descriptors were eliminated from the question. The translated questionnaire was pilot tested with eight young adults recruited from a university setting to assess face validity, with special focus on clarity, understandability, and ease of answering the questions. All translations were deemed clear, understandable, and relevant to the topic to these pilot participants. The final translated version was subsequently deployed for larger psychometric evaluation.

Phase 2: Psychometric Evaluation and Latent Class Analysis

Participants and settings

Associate and bachelor degree students between ages 18 and 24 were recruited from two major universities and their affiliate community colleges in Hong Kong. Flyers and mass emails with the study weblink to the online survey were distributed to students across campuses

between April and June of 2017. The survey may only be completed one time per electronic device to avoid multiple attempts by the same respondent. A total of 433 participants completed the Chinese ACE-IQ anonymously online by entering the study weblink; these participants may voluntarily provide their contact information if they agree to be contacted again for a study follow-up. The average age of participants in this full sample was 20.16 (SD =1.67); 178 were male (41.1%) and 218 were associate degree students (50.3%). Within three weeks of initial survey completion, a random sample of 32 participants who agreed to be contacted again completed the retest was 20.81 (SD =1.70); 8 were males (24.2%) and 13 were associate degree students (39.4%). In the re-test, Chinese and English versions of the ACE-IQ were administered in random order. In all instances of data collection, implied consent was obtained by way of survey completion.

Statistical analysis

All analyses were conducted using STATA SE14.1 (StataCorp, 2015). Descriptive statistics summarized participant characteristics. Test-rest reliability was examined by instrument item, domain subscale, and total score. Three reliability criteria were assessed. First, percentage agreement (PA) were calculated to assess the degree of absolute agreement between item response at test and retest; $PA \ge 70\%$ is considered satisfactory (Kazdin, 1977). Second, systematic differences in item-level responses at test and retest were further analyzed using the rank-invariant method (Svensson, 2012). The relative position (RP), i.e. the degree of difference in concentration of responses around central scale categories were calculated (Nordin, Murphy, & Danielsson, 2014; Svensson, 2012). Possible values of RP and RC range from -1 to 1;

higher absolute value implies higher degree of systematic disagreement between responses for an item across tests. A statistically significant systematic difference is noted when RP or RC confidence intervals do not straddle zero. Individual response invariability was assessed by relative rank variance (RV), which reflects the degree of heterogeneity in changes among individuals unaccounted by systematic disagreements (Svensson, 2012); RV<0.1 is generally regarded as negligible. Third, intra-class correlations (ICC) for test-retest reliability of the full instrument and its domain subscales were examined using 2-way mixed-effects models (DeVon et al., 2007; Koo & Li, 2016; McHugh, 2012). Evaluation criteria for ICC are as follows: ICC < 0.50 indicates poor reliability; ICC between 0.50–0.75 is considered moderate; ICC between 0.75–0.90 is good; and ICC > 0.90 is excellent. The statistical methods used to determine test-retest reliability were also employed to test the equivalence between Chinese and English versions of the questionnaire.

Patterns of childhood adversity were examined by classifying respondents into distinct groups based on their dichotomized responses (i.e. "exposed" and "not exposed") to the 13 ACE categories. LCA was conducted using the Penn State LCA Stata Plugin (The Methodology Center, 2015). The 13 variables were first entered into a 1-class model, and number of classes was incrementally increased until the best model was identified based on fit statistics, substantive interpretation, and proportion of participants represented in each class. Multiple model fit statistics were estimated to determine optimal number of classes, including Akaike's Information Criteria (AIC) (Akaike, 1987), Bayesian Information Criteria (BIC) (Schwarz, 1978), sample size adjusted BIC (ssaBIC)(Sclove, 1987), Entropy (Magidson & Vermunt, 2002), and p-value for bootstrap likelihood ratio tests (BSLRT) (Nylund, Asparouhov, & Muthén, 2007). Lastly, participants were assigned a pattern of ACE exposure using most likely class membership (Clark

& Muthén, 2009). Descriptive statistics summarized proportion of participants exposed to each ACE category and total number of ACEs for the full sample and by class. Differences in ACE score by class was examined using ANOVA with post hoc pairwise comparison using the Tukey test.

Results

Test-retest reliability

The time between initial test and retest ranged from 14 to 31 days (M = 22.2, SD = 6.2). Percentage agreement (PA), systematic disagreement (i.e., RP and RC), and individual variability (RV) between assessments of Chinese version at different time points are shown in **Table 1**. A satisfactory PA (\geq 70%) was noted for all questions, except Items P1-P2 (emotional neglect), A1-A2 (emotional abuse), and F6 (domestic violence). A small but significant systematic disagreement in relative position between test and retest was found for Item A3 (physical abuse). However, this item had satisfactory PA (81%) and the significant disagreement may be due to a minority of participants (n=4) who provided a very different response at retest (i.e. difference of \geq 2 on the 4-point Likert scale). Items A1-A2 (emotional abuse) and F6-F7 (domestic violence) also exhibited deviations in relative position of responses (i.e. lower exposure at retest), but they were not statistically significant. No differences in relative concentration and relative rank variance of responses were found between test and retest. The overall instrument demonstrated a good test-retest reliability (ICC=0.90); all three subscales also showed good test-retest reliability, with ICC ranging between 0.78 and 0.90.

Language equivalence

Semantic equivalence between Chinese and English versions of the ACE-IQ as measured in PA, RP, RC, and RV are also displayed in **Table 1.** Absolute agreement between items across languages was satisfactory overall; only two items had PA<70% (i.e. Items P2 and A1). A small but non-negligible systematic disagreement in relative position was found for Items P1-P2 (physical neglect). Item P2 also displayed a comparatively higher systematic disagreement in concentration (RC=0.12) and relative rank invariance (RV=0.20). The overall instrument demonstrated good equivalence across languages (ICC=0.90). ICC for "Household Dysfunction" and "Abuse and Neglect" domain subscales showed good equivalence, with ICC between 0.89 and 0.90; overall reliability of the "Community Violence" domain subscale across languages was moderate (ICC=0.63).

ACE exposure and latent class analysis

Descriptive statistics on ACE by type and cumulative exposure for the full sample are presented in **Table 2.** Nearly 3 out of 4 participants (74.36%) reported at least one ACE; nearly half (48.18%) reported two or more ACEs. The most commonly reported ACE was physical abuse (39.95%), followed by witnessing domestic violence (30.48%), parental death or separation (23.79%), and emotional abuse (20.32%). In latent class analysis, models with one to four classes were tested; the model fit statistics are displayed in **Table 3**. The 3-class model was selected based on lowest AIC and ssaBIC, and an insignificant BSLRT for the 4-class model indicating that the 3-class model was adequate for describing the data (Dziak, Lanza, & Tan, 2014). Additionally, even though the 4-class model had highest classification precision (i.e. entropy), one of the classes only represented 6 participants. Therefore, the 3-class model was deemed the best and most parsimonious solution.

The item-response probabilities and class prevalence for each class in the 3-class model are depicted in **Figure 1**. The largest class represent 65.82% of the participants and was labeled *Low ACEs* (Class 1). Overall, participants in this class had low probabilities of exposure to ACEs

across all categories (<0.20). The second largest class represent 24.94% of the participants, and was labeled *Household Violence* (Class 2). Participants in this class had high probabilities of endorsing exposure to physical abuse (0.96) and witnessing domestic violence in the household (0.67); they also had high probability of exposure to emotional abuse (0.59), but otherwise had low probabilities of endorsing other ACEs (<0.30). The final class was labeled *Multiple ACEs* (Class 3) and was represented by 9.24% of the participants. Participants in this class had high probabilities for endorsing all types of abuse (0.42-0.60), parental separation or death (0.62), and witnessing domestic violence (0.70). Additionally, they also had the highest probabilities for endorsing ACE categories related to family dysfunction (i.e. family member substance use, incarceration, and mental illness, parental separation or death, and witnessing domestic violence) and experiencing violence outside the home (i.e. bullying, community violence, and collective violence) compared with other classes.

Observed rates of ACEs exposure by category and total number for each class are also displayed in **Table 2**. Participants represented by *Multiple ACEs* reported highest proportions of exposure to all ACE categories compared with the other classes, except for emotional and physical abuse. Exposure rates by ACE category for each class and the full sample are presented in **Figure 2**. The total ACE score ranged from 0 to 4 for participants in *Low ACE*, 2 to 6 for *Household Violence*, and 3 to 7 or more for *Multiple ACEs*. Class membership was significantly associated with ACE score (*F*=519.71, *df*=2/430, *p*<0.001); Tukey post hoc analysis demonstrated that all pairwise comparisons were statistically significant (*p*<0.001), with participants in the *Low ACE* class reporting lowest mean ACE score (mean=0.83, SD=0.82) followed by *Household Violence* (mean=3.31, SD=0.94) and *Multiple ACEs* (mean=4.95, SD=1.53).

Discussion

The present study provides the first Chinese translation and psychometric evaluation of the ACE-IQ using a sample of 433 Chinese young adults. This study also produced novel evidence of different patterns of ACE exposure based on a full constellation of 13 ACEs and outside a Western context. Our findings showed that the Chinese ACE-IQ has sufficient scale reliability overall, and good content validity and semantic equivalence with the original English version. Three patterns of ACE exposure, with the largest group under *Low ACEs* (66%) followed by *Household Violence* (25%) and *Multiple ACEs* (9%) will be discussed.

The test-retest reliability of the Chinese ACE-IQ for the full scale and by domain subscales was good. Absolute agreement between test-retest by item was generally satisfactory (i.e. majority above 70%), except for questions related to emotional neglect, emotional abuse, and aggression between household members. This may be due to inherent difficulties in defining psychological maltreatment (Glaser, 2002), particularly across cultural norms that vary in views and knowledge on child safety and protection, and suggests experiences related to psychological aggression in childhood may be more amenable to recall bias. However, these results should also be interpreted with caution given the small and atypical sample of retest participants; more research is needed to assess the test-retest reliability of individual items, particularly those related to childhood maltreatment, using larger representative samples. Similarly, semantic equivalence by domain subscales was generally satisfactory. Of note, overall language equivalence of the "Community Violence" subscale was moderate, which may be due to the use of unfamiliar collective violence terms and their rare occurrence in Hong Kong. We acknowledge the low relevance of some of these items to the local context, but have retained them for future use and comparisons. The current study sets the stage for continued research using the ACE-IQ to assess ACE exposure in other Chinese populations.

Nearly 75% of the participants reported at least one ACE: 31% reported three or more ACEs. This finding suggests that the overall prevalence of ACEs reported by Hong Kong college students was higher compared with international norms. For example, according to the World Mental Health survey of 51,945 adults across 21 countries, 39% of participants reported at least one ACE and only 6% reported three or more ACEs (Ronald C Kessler et al., 2010). However, it is possible that these differences stemmed from how ACEs were measured across studies. Specifically, the World Mental Health survey included some additional ACEs (e.g. lifethreatening illness and economic hardship) while excluding others (e.g. emotional abuse and neglect, bullying, and community/ collective violence) that were included in the present study. Further, the World Mental Health survey measured ACEs using direct interviews, while the present study used anonymous web-based surveys. It is possible that different data collection methods (e.g. face-to-face interviews versus anonymous response) can influence individuals' reports of their childhood adversities. Indeed, we identified two studies that were conducted in an Asian country that employed anonymous self-reports and used the same adjusted frequency scoring method of the ACE-IO. Our results are similar to those of a study of 2,099 university students across eight provinces in Vietnam, where 76% reported at least one ACE and 37% reported three or more ACEs (Quynh Anh Tran, Michael P Dunne, Thang Van Vo, & Ngoc Hoat Luu, 2015). Another study of 939 Korean college students showed a comparatively lower rate of ACE exposure, where 50% reported at least one ACE and 15% reported three or more ACEs (Kim, 2017).

Three distinct patterns of ACE exposure were identified among our sample of Chinese young adults, two of which have been described extensively in prior studies - Low ACEs and Multiple ACEs (Cavanaugh et al., 2015; Leslie E Roos et al., 2016; Ross, Waterhouse-Bradley, Contractor, & Armour, 2018; Shin et al., 2018). Our results are generally consistent with previous studies suggesting that about two-thirds of young adults have low exposure to ACEs and about one in ten experienced multiple types of childhood adversities. However, contrary to prior findings from the US or Canada, we did not find a distinct pattern of exposure related exclusively to household dysfunction (Cavanaugh et al., 2015; L. E. Roos et al., 2016; Shin et al., 2018). Instead, we uncovered a novel pattern of exposure involving high levels of violence within the home, both directed towards the child and towards other household members. Young adults represented by the Household Violence class experienced the highest concurrent exposure of physical abuse, domestic violence, and emotional abuse. This may reflect a combination of Chinese traditional norms, including rigid gender roles, absolute parental authority, and endorsement of physical punishment (Chan, 2009; Ho & Gross, 2015; Zhai & Gao, 2009), that may potentiate risks for family violence in the absence of other household risk factors (e.g. substance use or mental illness). Indeed, a survey of 1,094 school-aged children in Hong Kong demonstrated that the rate of co-occurrence of child maltreatment and domestic violence is high, where over 50% of maltreated children also witnessed domestic violence, with lifetime cooccurrence prevalence of 18% among all children (Chan, 2011). Our results highlight the importance of examining ACE exposure within local contexts as children's adverse experiences can be idiosyncratic to geographic, social, and cultural norms. Future studies may also include representative samples of Chinese respondents across the age spectrum to explore possible

historical effects on how ACEs may occur or co-occur differently over time under this cultural context.

A significant limitation of this study is the convenient and age-restricted nature of the participant sample. Our sample of young college students from Hong Kong may not fully reflect Chinese populations across geographic and cultural contexts, thereby limiting the generalizability of our findings. Although young adults may more accurately recall childhood experiences, replication with older and more representative samples across different settings is needed. Also, our retest sample was small and characteristically different from the full sample. Thus, further reliability testing of the Chinese ACE-IQ is warranted. Lastly, as discussed above, some of the items retained in the Chinese ACE-IQ may not be applicable to the experiences of the study participants. However, they were retained to create a complete measure that may be applied in different Chinese populations for future research.

ACEs are now widely recognized as a key source of social and health disparity that can significantly alter children's life trajectories (Shonkoff et al., 2012). Therefore, early identification and interventions geared towards addressing the potential downstream outcomes of childhood adversities are critical to anticipating client needs and improving overall population health. Routine ACE screening within primary care settings may be the key to propelling these efforts, and some evidence shows that these measures are both feasible and effective for identifying at-risk populations (Glowa, Olson, & Johnson, 2016). Our findings highlight the importance of screening ACEs among Hong Kong young adults given their comparatively higher rates of exposure. At a minimum, our findings support routine screening of physical abuse, emotional abuse, and domestic violence given the novel pattern of exposure found in this local context.

Emerging research demonstrates that providing safe and nurturing childhood environments to support vulnerable children and families can effectively reverse or mitigate the negative outcomes associated with ACEs (Bellis et al., 2017). Therefore, prevention and intervention strategies that reduce children's direct and indirect exposure to aggression and violence in the home are necessary. Our results call for continued investigations into the design, evaluation, and effectively implementation of health and public policy initiatives that consolidate efforts across sectors to support this cause. Lastly, a broader examination of traditional cultural ideologies and how they may serve as risk or protective factors for ACEs across other Chinese populations are also warranted.

As a response to the need to develop a global ACEs surveillance framework (Anda et al., 2010), this study provides the first translation and validation of the ACE-IQ that can be employed in Chinese populations. Further, a novel pattern of ACE exposure was identified that may be specific to the Chinese and other non-Western cultures. This finding underscores the need to examine ACE exposure within local contexts, and call for further investigations on how culture-specific patterns of ACE exposure may confer different risks for the myriad of negative outcomes associated with childhood adversities.

References

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52(3), 317-332.

- Anda, R. F., Butchart, A., Felitti, V. J., & Brown, D. W. (2010). Building a framework for global surveillance of the public health implications of adverse childhood experiences. *American Journal of Preventive Medicine*, 39(1), 93-98. doi:10.1016/j.amepre.2010.03.015
- Bellis, M. A., Hardcastle, K., Ford, K., Hughes, K., Ashton, K., Quigg, Z., & Butler, N. (2017).
 Does continuous trusted adult support in childhood impart life-course resilience against adverse childhood experiences a retrospective study on adult health-harming behaviours and mental well-being. *BMC Psychiatry*, 17. doi:10.1186/s12888-017-1260-z
- Cavanaugh, C. E., Petras, H., & Martins, S. S. (2015). Gender-specific profiles of adverse childhood experiences, past year mental and substance use disorders, and their associations among a national sample of adults in the United States. *Social Psychiatry and Psychiatric Epidemiology*, 50(8), 1257-1266. doi:10.1007/s00127-015-1024-3
- Chan, K. L. (2009). Protection of face and avoidance of responsibility: Chinese men's account of violence against women. *Journal of Social Work Practice*, *23*(1), 93-108.
- Chan, K. L. (2011). Children exposed to child maltreatment and intimate partner violence: A study of co-occurrence among Hong Kong Chinese families. *Child Abuse and Neglect*, 35(7), 532-542. doi:<u>https://doi.org/10.1016/j.chiabu.2011.03.008</u>
- Chan, K. L. (2012). Comparison of Parent and Child Reports on Child Maltreatment in a Representative Household Sample in Hong Kong. *Journal of Family Violence*, 27(1), 11-21. doi:10.1007/s10896-011-9405-1
- Chan, K. L. (2013). Victimization and poly-victimization among school-aged Chinese adolescents: prevalence and associations with health. *Preventive Medicine*, *56*(3-4), 207-210.

- Clark, S. L., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis.
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., . . . Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*, 39(2), 155-164.
- Ding, Y., Lin, H., Zhou, L., Yan, H., & He, N. (2014). Adverse childhood experiences and interaction with methamphetamine use frequency in the risk of methamphetamine-associated psychosis. *Drug and Alcohol Dependence, 142*, 295-300.
- Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect Size, Statistical Power and Sample Size
 Requirements for the Bootstrap Likelihood Ratio Test in Latent Class Analysis. *Structural* equation modeling : a multidisciplinary journal, 21(4), 534-552.
 doi:10.1080/10705511.2014.919819
- Fang, X., Fry, D. A., Ji, K., Finkelhor, D., Chen, J., Lannen, P., & Dunne, M. P. (2015). The burden of child maltreatment in China: a systematic review. *Bulletin of the World Health Organization*, 93(3), 176-185C.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . .
 Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, 14(4), 245-258.
- Glaser, D. (2002). Emotional abuse and neglect (psychological maltreatment): A conceptual framework. *Child abuse & neglect*, *26*(6-7), 697-714.

- Glowa, P. T., Olson, A. L., & Johnson, D. J. (2016). Screening for adverse childhood experiences in a family medicine setting: a feasibility study. *The Journal of the American Board of Family Medicine*, 29(3), 303-307.
- Ho, G. W., & Gross, D. A. (2015). Differentiating physical discipline from abuse: Q findings from Chinese American mothers and pediatric nurses. *Child Abuse and Neglect, 43*, 83-94.
- Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis*, 10(1), 141-150.
- Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A.
 M., . . . Williams, D. R. (2010). Childhood adversities and adult psychopathology in the
 WHO World Mental Health Surveys. *British Journal of Psychiatry*, 197(5), 378-385.
- Kim, Y. H. (2017). Associations of adverse childhood experiences with depression and alcohol abuse among Korean college students. *Child Abuse and Neglect*, *67*, 338-348.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.
- Lee, S., Guo, W., Tsang, A., He, Y., Huang, Y., Zhang, M., . . . Kessler, R. (2011). The prevalence of family childhood adversities and their association with first onset of DSM-IV disorders in metropolitan China. *Psychological Medicine*, *41*(1), 85-96.
- Li, Y., Cao, F., Cao, D., & Liu, J. (2015). Nursing students' post-traumatic growth, emotional intelligence and psychological resilience. *Journal of Psychiatric and Mental Health Nursing*, 22(5), 326-332.
- Liu, Z., Yang, Y., Shi, Z., Liu, J., & Wang, Y. (2016). The risk of male adult alcohol dependence: The role of the adverse childhood experiences and ecological executive function. *Comprehensive Psychiatry*, 68, 129-133.

- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*, *35*(6), 382-386.
- Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with Kmeans. *Canadian Journal of Marketing Research*, 20(1), 36-43.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282.
- Nordin, Å., Murphy, M. A., & Danielsson, A. (2014). Intra-rater and inter-rater reliability at the item level of the Action Research Arm Test for patients with stroke. *Journal of rehabilitation medicine*, *46*(8), 738-745.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, 14(4), 535-569.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health, 29*(5), 489-497.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, *30*(4), 459-467.
- Roos, L. E., Afifi, T. O., Martin, C. G., Pietrzak, R. H., Tsai, J., & Sareen, J. (2016). Linking typologies of childhood adversity to adult incarceration: Findings from a nationally representative sample. *American journal of orthopsychiatry*, *86*(5), 584.
- Ross, J., Waterhouse-Bradley, B., Contractor, A. A., & Armour, C. (2018). Typologies of adverse childhood experiences and their relationship to incarceration in US military veterans. *Child abuse & neglect, 79*, 74-84.

- Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343. doi:10.1007/bf02294360
- Shin, S. H., McDonald, S. E., & Conley, D. (2018). Patterns of adverse childhood experiences and substance use among young adults: a latent class analysis. *Addictive behaviors*, 78, 187-192.
- Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., Garner, A. S., . . . Wood, D. L. (2012). The Lifelong Effects of Early Childhood Adversity and Toxic Stress. *Pediatrics*, 129(1), e232-e246. doi:10.1542/peds.2011-2663

StataCorp. (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

- Stoltenborgh, M., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., & Alink, L. R. A. (2013). Cultural–geographical differences in the occurrence of child physical abuse? A metaanalysis of global prevalence. *International Journal of Psychology*, 48(2), 81-94. doi:10.1080/00207594.2012.697165
- Svensson, E. (2012). Different ranking approaches defining association and agreement measures of paired ordinal data. *Statistics in Medicine*, *31*(26), 3104-3117.
- The Methodology Center. (2015). LCA Stata Plugin (Version 1.2) [Software]. University Park: Penn State. Retrieved from methodology.psu.edu
- Tran, Q. A., Dunne, M. P., Vo, T. V., & Luu, N. H. (2015). Adverse childhood experiences and the health of university students in eight provinces of Vietnam. *Asia Pacific Journal of Public Health*, 27(8_suppl), 26S-32S.
- Wei, W. (2013). Adverse childhood experience as a risk factor for adolescent pregnancy in China. International Journal of Child and Adolescent Health, 6(3), 323.

- Wong, J. Y.-H., Fong, D. Y.-T., & Chan, K. K.-W. (2015). Anxiety and insomnia as modifiable risk factors for somatic symptoms in Chinese: a general population-based study. *Quality of Life Research*, 24(10), 2493-2498.
- World Health Organization [WHO]. (2016). Adverse Childhood Experiences International Questionnaire (ACE-IQ). Retrieved from

http://www.who.int/violence_injury_prevention/violence/activities/adverse_childhood_exper iences/en/

- Xiao, Q., Dong, M.-X., Yao, J., Li, W.-X., & Ye, D.-Q. (2008). Parental Alcoholism, Adverse Childhood Experiences, and Later Risk of Personal Alcohol Abuse among Chinese Medical Students. *Biomedical and Environmental Sciences*, *21*(5), 411-419. doi:https://doi.org/10.1016/S0895-3988(08)60062-8
- Xue, J.-M., Gao, W., & Cao, F.-L. (2016). Quality of life among breast cancer survivors 2 years after diagnosis: the relationship with adverse childhood events and posttraumatic growth. *Cancer Nursing*, 39(3), E32-E39.
- Xue, J.-M., Lin, P.-Z., Sun, J.-W., & Cao, F.-L. (2017). Disrupted executive function and aggression in individuals with a history of adverse childhood experiences: an event-related potential study. *The Journal of nervous and mental disease*, *205*(12), 942-951.
- Zhai, F., & Gao, Q. (2009). Child Maltreatment Among Asian Americans: Characteristics and Explanatory Framework. *Child Maltreatment*, 14(2), 207-224.
 doi:10.1177/1077559508326286

Running head: Chinese ACE-IQ

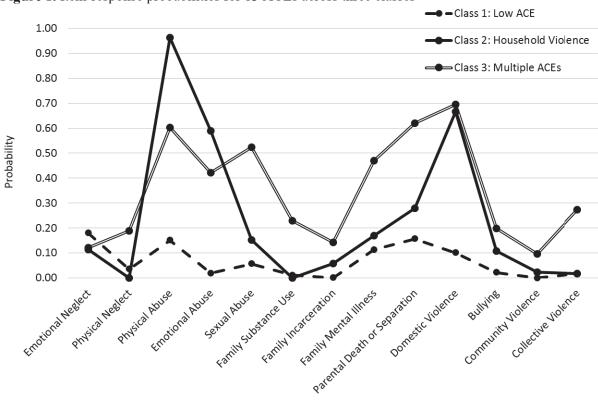
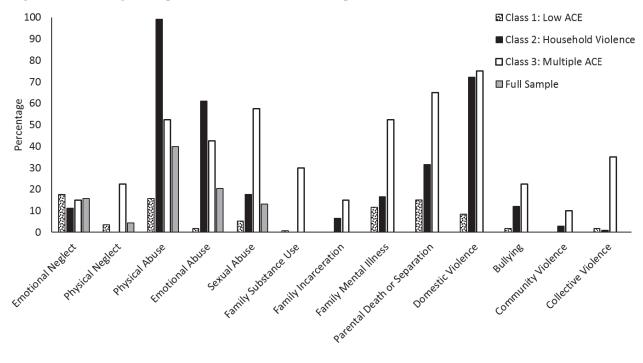


Figure 1. Item-response probabilities for 13 ACEs across three classes

Figure 2. Percentage of exposure to 13 ACEs in full sample and across the three classes



Running head: Chinese ACE-IQ

Table 1. Test-retest reliability and semantic equivalence of the Chinese ACE-IQ by item, domai	n subscale,
and total score	

		Т	est-Retes	t Reliabi	lity	Semantic Equivalence			
#	Item	PA	RP	RC	RV	РА	RP	RC	RV
Chilo	lhood Maltreatment	1	<u> </u>					<u> </u>	1
P1	Did your parents/guardians understand your problems and worries?	63	-0.01	0.08	0.02	75	0.12*	0.01	0.02
P2	Did your parents/guardians really know what you were doing with your free time when you were not at school or work?	66	0	-0.07	0.09	66	0.26*	0.12	0.20
P3	How often did your parents/guardians not give you enough food even when they could easily have done so?	91	-0.01	0.02	0.01	94	0.01	-0.02	0.01
P4	Were your parents/guardians too drunk or intoxicated by drugs to take care of you?	97	-0.03	0	0	100	0	0	0
P5	How often did your parents/guardians not send you to school even when it was available?	97	-0.03	0	0	91	0.09	0	0
Al	Did a parent, guardian or other household member yell, scream or swear at you, insult or humiliate you?	63	-0.13	-0.05	0.03	69	0.01	0.11	0.04
A2	Did a parent, guardian or other household member threaten to, or actually, abandon you or throw you out of the house?	69	-0.18	-0.06	0.01	100	0	0	0
A3	Did a parent, guardian or other household member spank, slap, kick, punch or beat you up?	81	-0.15*	0.04	0.02	88	0	0	0.01
A4	Did a parent, guardian or other household member hit or cut you with an object, such as a stick (or cane), bottle, club, knife, whip etc?	74	-0.04	0.03	0.03	94	-0.03	-0.03	0.01
A5	Did someone touch or fondle you in a sexual way when you did not want them to?	88	-0.10	0.07	0.01	94	0.04	-0.07	0.01
A6	Did someone make you touch their body in a sexual way when you did not want them to?	100	0	0	0	97	0.03	-0.03	0.01
A7	Did someone attempt oral, anal, or vaginal intercourse with you when you did not want them to?	100	0	0	0	97	0.03	-0.03	0.01

A8	Did someone actually have oral, anal, or vaginal intercourse with you when you did not want them to?	100	0	0	0	97	0.03	-0.03	0.01
	Domain Subscale ICC:		0.89 (0.	78-0.94)			0.89 (0.7	78-0.94)	
Fami	ly/ Household Dysfunction	1				I			
F1	Did you live with a household member who was a problem drinker or alcoholic, or misused street or prescription drugs?	97	0.03	0	0	100	0	0	0
F2	Did you live with a household member who was depressed, mentally ill or suicidal?	94	0.04	0	0.01	94	0.01	0	0.01
F3	Did you live with a household member who was ever sent to jail or prison?	97	-0.03	0	0	100	0	0	0
F4	Were your parents ever separated or divorced?	97	0.03	0	0	100	0	0	0
F5	Did your mother, father or guardian die?	92	0.03	0	0	100	0	0	0
F6	Did you see or hear a parent or household member in your home being yelled at, screamed at, sworn at, insulted or humiliated?	53	-0.11	-0.01	0.06	72	-0.01	0.05	0.07
F7	Did you see or hear a parent or household member in your home being slapped, kicked, punched or beaten up?	84	-0.12	-0.05	0.02	81	0.04	0.1	0.01
F8	Did you see or hear a parent or household member in your home being hit or cut with an object, such as a stick (or cane), bottle, club, knife, whip etc.?	81	0.06	0.01	0.01	91	-0.04	0.04	0
	Domain Subscale ICC:		0.90 (0.	80-0.95)			0.90 (0.8	80-0.95)	1
Viole	nce Outside the Home	1				<u> </u>			
V1	How often were you bullied?	75	0.02	0.12	0.02	97	-0.03	-0.01	0.01
V4	Did you see or hear someone being beaten up in real life?	72	-0.03	0.13	0.01	72	-0.11	0.09	0.05
V5	Did you see or hear someone being stabbed or shot in real life?	91	-0.03	0.06	0.01	100	0	0	0
V6	Did you see or hear someone being threatened with a knife or gun in real life?	94	-0.01	0.03	0.01	94	0	0	0.01
V7	Were you forced to go and live in another place due to any of these events?	100	0	0	0	97	0.03	0	0
V8	Did you experience the deliberate	100	0	0	0	100	0	0	0

	destruction of your home due to any of these events?								
V9	Were you beaten up by soldiers, police, militia, or gangs?	97	0.03	0	0	100	0	0	0
V10	Was a family member or friend killed or beaten up by soldiers, police, militia, or gangs?	94	-0.06	-0.03	0	100	0	0	0
	Domain Subscale ICC		0.78 (0.	54-0.89)			0.63 (0.3	37-0.80)	
Full S	scale ICC		0.90 (0.	81-0.95)		0.90 (0.81-0.95)			

Note: PA = percentage agreement; RP = relative position; RC = relative concentration; RV = relative rank variance; ICC = intra-class correlation; * denotes statistically significant systematic difference; Evaluation reliability criteria for ICC: poor = ICC < 0.50; moderate = ICC of 0.50–0.75; good = ICC of 0.75–0.90; and excellent = ICC > 0.90.

Running head: Chinese ACE-IQ

	FullClass 1 :Class 2:Class					
	Sample	Low ACE	Household Violence	Multiple ACEs		
	(n=433)	(n=285)	(n=108)	(n=40)		
	%	%	%	%		
Emotional Neglect	15.70	17.54	11.11	15.00		
Physical Neglect	4.39	3.51	0.00	22.50		
Physical Abuse	39.95	15.79	99.07	52.50		
Emotional Abuse	20.32	1.75	61.11	42.50		
Sexual Abuse	13.16	5.26	17.59	57.50		
Family Substance Use	3.23	0.70	0.00	30.00		
Family Incarceration	3.00	0.00	6.48	15.00		
Family Mental Illness	16.63	11.58	16.67	52.50		
Parental Death or Separation	23.79	15.09	31.48	65.00		
Domestic Violence	30.48	8.42	72.22	75.00		
Bullying	3.24	1.75	12.04	22.50		
Community Violence	1.62	0.00	2.78	10.00		
Collective Violence	4.62	1.75	0.93	35.00		
Number of ACEs						
0	25.64	38.95	0	0		
1	28.18	42.81	0	0		
2	15.01	14.74	21.30	0		
3	12.47	3.16	36.11	15.00		
4	11.78	0.35	34.26	32.50		
5	3.46	0	6.48	20.00		
6	2.31	0	1.85	20.00		
7 or more	1.15	0	0	12.50		
Mean Number of ACEs (SD)*	1.83 (1.73)	0.83 (0.82)	3.31 (0.94)	4.95 (1.53)		

Table 2. ACEs	prevalence rates b	y category and total	l number for full sam	ple and by class
---------------	--------------------	----------------------	-----------------------	------------------

* Significant overall difference across 3 classes and for all pairwise comparisons in post hoc analysi (*p*<0.001).

Classes	Log likelihood	AIC	BIC	ssaBIC	Entropy	BSLRT p-value
1	-1981.54	815.72	868.64	827.36		
2	-1833.57	547.78	657.69	572.01	0.75	0.01
3	-1807.78	524.21	691.11	561.00	0.78	0.01
4	-1800.60	537.85	761.74	587.20	0.83	1.0

Table 3. Fit indices for l	latent class analysis	for 1-5 models
----------------------------	-----------------------	----------------

Note: AIC = Akaike information criterion; BIC = Bayesian information criterion; ssaBIC = Sample-size adjusted BIC; BSLRT = Bootstrapped likelihood ratio test; Final selected model is in **bold**.