Can well-being be measured using Facebook status updates? Validation of Facebook's Gross National Happiness Index

Wang, N., Psychometrics Centre, University of Cambridge, UK

Kosinski, M., Psychometrics Centre, University of Cambridge, UK

Stillwell, D. J., School of Psychology, University of Nottingham, UK

Rust, J., Psychometrics Centre, University of Cambridge, UK

Abstract:

Facebook's Gross National Happiness (FGNH) indexes the positive and negative words used in the millions of status updates submitted daily by Facebook users. FGNH has face validity: it shows a weekly cycle and increases on national holidays. Also, happier individuals use more positive words and fewer negative words in their status updates (Kramer, 2010). We examined the validity of FGNH in measuring mood and well-being by comparing it with scores on Diener's Satisfaction with Life Scale (SWLS), administered to an average of 34 Facebook users every day for a year, then aggregated by day, week, month, quarter and half year. FGNH and SWLS were not significantly correlated, with a negative correlation coefficient. Also, aggregated SWLS scores showed a positive relationship with numbers of negative words in status updates. We conclude that FGNH is a valid measure for neither mood nor well-being; however, it may play a role in mood regulation. This challenges the assumption that linguistic analysis of internet messages is related to underlying psychological states.

Article text:

Introduction

The pervasiveness of digital communication and especially the rise of social media such as blogs, has created an opportunity and spurred interest in automated sentiment analysis – computer analysis of digital texts aimed at extracting authors' emotions or attitudes (Nasukawa & Yi, 2003). Progress in computational methods such as Machine Learning and Latent Semantic Analysis, and quickly growing processing power allows the application of sentiment analysis on an unprecedented scale. It has been used in numerous realms, such as understanding customers' feedback, investigating political discourse, opinion mining, and tracking changes in individual and group mood (Gamon, 2004; Mishne & Rijke, 2006; Mullen & Malouf, 2006; Pang & Lee, 2008).

Recently, the Facebook Data Team released a novel metric based on automated sentiment analysis called Facebook Gross National Happiness (FGNH; Facebook Data Team, 2011). The concept of Gross National Happiness (GNH) was first proposed in 1972 by the former King of Bhutan, Jigme Singye Wangchuck, as part of his attempt to develop the country according to its Buddhist culture (Bates, 2009; Pennock, 2009; Pennock & Ura, 2011). The concept was later adopted by western countries, where it has been argued that GNH is a better indicator of quality of life than the single value Gross Domestic Product. GNH embraces both physical and mental health, being influenced by a combination of education, ecological vitality, living standards, good governance, community vitality, physical health and psychological well-being (Pennock & Ura, 2011). In the psychology literature GHN has usually been related to "well-being" rather than to its literal meaning of 'an emotional state of pleasure' (Pavot & Diener, 2009). Indeed, the UK-based Sustainable Development Research Network regards the concepts of "GNH", "quality of life”, “life satisfaction” and ”well-being”, as synonymous on most occasions (McAllister, 2005).

While GNH is easy to understand, it is hard to measure, and to date no method of measurement has been globally accepted. Most of the current measurements are either still under development and need improvement, or borrow from existing psychometric tests in related realms. Furthermore, almost all current GNH measurements are indirect and are operationally defined as combinations of factors that emerge from pre-existing scales, hence their validity as measures of GNH itself is unclear (Institute of Wellbeing, 2009; Pennock, 2009; Pennock & Ura, 2011; Ura, 2008)

Facebook’s GNH measure is an attempt to estimate the aggregated mood and well-being of the Facebook population, applying automated sentiment analysis to the status updates of millions of Facebook users. Status updates are short-format notes (the average number of words of each status update is nine) that can be seen by some or all of a Facebook user’s friends (Kramer, 2010). The content of status updates usually involves moods, ideas, events and states recently encountered in the users’ lives, as well as messages they want to broadcast to their friends (e.g. “*Happy New Year to you all*”). FGNH is based on the assumption that the more positive words people use in their status updates on a certain day, the happier they are, and vice versa (Kramer, 2010). FGNH offers three indexes: *positivity* is the number of positive words in users’ status updates, *negativity* is the number of negative words, and the main index is the standardized difference between positivity and negativity.

To identify and count the positive and negative words appearing in status updates, the Facebook Data Team employed the Text Analysis and Word Count (TAWC) program, which is built upon the Linguistic Inquiry and Word Count 2007 (LIWC2007) dictionary. It contains around 4500 words and word stems, grouped into 64 categories. The category ‘positive emotion’ has 407 words or word stems in total, including words like *“love”,” nice”,* and *“sweet”*. The ‘negative emotions’ category, with 506 words or word stems, includes words like “*hurt”, “ugly”,* and “*nasty”*. It has been argued that the LIWC2007 is psychologically and psychometrically valid (Pennebaker, Chung, Ireland, A Gonzales, & Booth, 2007; Tausczik & Pennebaker, 2010).

FGNH index, as well as its two components Positivity and Negativity, is available through the “Gross National Happiness” Facebook application (Facebook Data Team, 2011). Even though FGNH was first published in 2009, the index reaches back as far as September 2007 (Kramer, 2010). In 2010 the FGNH was extended to include four more languages (Dutch, German, Italian and Spanish) and 18 more countries (Facebook Data Team, 2010).

As a practical measure of aggregated mood, FGNH has much in its favor. It is based on the status updates of hundreds of millions of people, and yet it is collected automatically each day from incidental information without further cost. The format of updates is brief, compared to blogs or other forms of digital communication, making emotional words easily recognizable.

Additionally, the self-descriptive property of status updates seems to encourage users to express their feelings and embed emotional contents. Moreover, most updates are not targeted at a particular individual, hence they elicit emotional rather than relationship information (Kramer, 2010).

It has been suggested that FGNH can be interpreted on two levels. Its daily fluctuations might indicate the short-term changes in the mood of the Facebook population. Kramer (2010) pointed out that the peaks in the FGNH graph (see Figure 1) appear during events commonly believed to cheer people up, such as Thanksgiving and Christmas, while troughs appear when depressing events or commemoration days occur, such as after Michael Jackson’s sudden death (Kramer, 2010). Additionally, FGNH follows a weekly cycle with a peak on Friday and a drop on Monday, which fits well the intuitive weekly cycle of happiness. On the other level, long-term changes in the FGNH (e.g. monthly or yearly) are believed to express the change in an aggregated wellbeing of the Facebook population. Kramer (2010) shows that there is a modest but significant positive correlation (r=0.17, p< 0.001) between Satisfaction with Life scores (SWLS, Diener et al 1984) and the overall positivity of individuals' status updates (calculated by a similar formula to FGNH, but here they excluded positive words in terms like "Happy New Year"). In other words, happy people tend to use more positive words in their status updates.

In the current study, we examined the validity of the FGNH as a measure of both mood and well-being. This is an important topic because researchers are beginning to use and draw conclusions from linguistic measures of mood based on other large online datasets such as from Twitter (Golder & Mach, 2011) without confirming a relationship with self-reported mood. We used a large sample of Satisfaction with Life scores acquired in 2009 that are known to be composed of relatively stable well-being and momentary mood. We aggregated SWLS scores by days, weeks and months in order to establish a benchmark of mood and well-being fluctuations in the Facebook population. We then compared these values with FGNH indexes in order to examine FGNH validity.

Method

Similarly to Kramer (2010) we adopted the Satisfaction With Life Scale questionnaire (Diener, Emmons, Larsen, & Griffin, 1985) as a proxy for happiness. The SWLS is a widely used measure of global life satisfaction, composed of 5 items with a Likert response scale. In our study, the SWLS scale showed a high reliability, as indicated by a Cronbach's Alpha coefficient of 0.82.

The SWLS was shown to be a valid metric for an individual's subjective judgment of his/her own global life satisfaction level, based on endogenous criteria (Diener, 1984; Diener, Emmons, Larsen, & Griffin, 1985). However, SWL was also shown to be influenced by momentary judgment based on short term cues accessible around the test taking time (Schimmack, Diener, & Oishi, 2002; Schwarz & Strack, 2003). Consequently, SWLS scores can shift as a function of momentary mood but its average level is anchored at a more permanent level of general happiness. In this study, we use daily fluctuations of aggregated SWLS scores as a proxy for short-term changes in users' mood, while SWLS scores aggregated by weeks and months served as a proxy of changes in general well-being.

The dataset used in this study was collected by the *myPersonality* Facebook application, which offers genuine personality assessments to Facebook users (Stillwell & Kosinski, 2011). In return for their time and effort, participants were provided with personalized feedback on their SWLS scores. The SWLS was published on myPersonality in July 2008. So far, it has been taken more than 120,000 times. In this study we selected the SWLS scores collected in year 2009 from US Facebook users, aged 16 to 60. To make the aggregated SWLS scores reliable, we excluded days with fewer than 10 individual scores (29 days). The number of respondents per day ranged from 10 to 94 (M =34.10, SD=17.88). There were 24,193 users in our sample (58% females, 9% unidentified; average age M =25.43, SD=7.90).

Daily values of Facebook Gross National Happiness (Facebook Data Team, 2011) together with Positivity and Negativity indexes were downloaded from the Facebook Gross National Happiness application (Facebook Data Team, 2011). FGNH is operationalized as the difference between Positivity and Negativity indexes and indicates amount of positive versus negative words used in status updates on a given day. All measures used in this research were standardized into Z-scores before the aggregation.

Results

Results of this study indicate that Facebook Gross National Happiness does not have convergent validity with the Satisfaction with Life Scale, and thus might not be a valid measure of mood and well-being. Moreover, it seems that FGNH might be negatively related to happiness and general well-being.

The Pearson correlations between FGNH and SWLS on a daily, weekly, and monthly level are presented in Table 1. On a daily and weekly levels, SWLS correlates positively ($r=.13$ and $r=.37$ respectively) with Negativity. On a monthly level positive correlation with Negativity becomes relatively strong ($r=.72$), while negative correlation with FGNH ($r=-.64$) becomes significant.

Figure 2 presents the monthly changes in standardized SWLS, FGNH, Positivity, and Negativity values. Consistent with the correlation values, SWLS scores aggregated by month follow changes in Negativity.

The negative relationship between SWLS and FGNH becomes more apparent on the quarter and half year level (Figures 3 and 4). In most periods, when FGNH is above the mean, SWLS is below, and vice versa.

Additional analyses were carried out to ascertain that the negative relationship between FGNH and SWL was not introduced by the fact that on certain occasions people may tend to use increased numbers of positive or negative words in their status updates regardless of their actual mood (e.g. on the New Year's Eve). In the subsample of 19 especially happy and especially unhappy days (above or below 1 SD from the mean) the correlations between SWLS and FGNH indexes were non-significant ($r_{SWLS\ \&\ FGNH}=-0.078$, $p=0.751$; $r_{SWLS\ \&\ POS}=-0.081$, $p=0.741$; r_{SWLS}

$_{\& NEG}$ =0.052, p=0.833) but similar to the results acquired from the entire sample. This suggests that the relationship between SWLS and FGNH is consistent across days characterized by both regular and outstanding numbers of positive and negative words used in status updates.

Further analyses were carried out to investigate the possibility that the negative but not significant correlation between FGNH and SWLS could be introduced by a sampling bias. For instance, it is possible that unhappy individuals were more likely to participate in the SWLS survey on days characterized by high FGNH. Two strategies were adopted to rule out sample bias. First, we examined the distribution of standard deviations (*SD*s) for measures used in this study across days. If a sample leans heavily to one extreme of the scale, floor or ceiling effects will lead to decreased *SD*. In general, we found that there was no relationship between the SWLS's *SD* and FGNH by day (r = - 0.01, p>0.05). A similar result was obtained using weekly aggregates (r= - 0.20, ns).

Second, we investigated the aggregated personality and demographic profiles of participants under the assumption that the potential sample bias is unlikely to affect SWLS scores exclusively, and should be visible in the fluctuations of the personality, age, and gender structure of the sample. We used age and gender recorded in respondents' Facebook profiles and their scores on the 100 item IPIP proxy for the NEO big five personality questionnaire (Goldberg, et al., 2006) stored in the myPersonality database. The Analysis of Variance showed no significant differences in the personality, age and gender structure of our sample between happy (FGNH higher than 1 *SD* above the mean) and regular days (F=0.03, ns). The two analyses mentioned above allowed us to conclude that the structure of the SWLS sample used in this study was unlikely to fluctuate with changing levels of FGNH index.

Discussion

Our results suggest that FGNH does not provide a true reflection of people's mood on a given day, nor is related to general well-being. If anything, it seems that the number of negative words used in Facebook status updates constitutes a positive indicator of good mood and well-being. Surprising as they are, our results might have several possible explanations. First, the relationship between Facebook status updates and general mood or well-being of Facebook users might not be straightforward. People may try to disguise their real emotions to present a contrived image to their Facebook friends. Second, users' statuses might be to some extent misinterpreted by the sentiment analysis algorithm that might be insensitive to individual habits of language use. For instance, the sentence "*I am not happy today*" may be interpreted as positive due to the word "happy", while *"I am unhappy"* would be treated as negative, even though they have the same meaning. Also, the context of users' updates might be hard to grasp. '*Happy Mother's Day'* does not carry the same salience as *'I'm extremely happy today'*. Similarly, semantic analysis software might be prone to misinterpreting ironic expressions, for example "*Today I was fired. Great*." An additional consideration is that the LIWC2007 dictionary was designed for traditional forms of written language, which may be unsuitable for the language used in the on-line environment. For example, it does not contain smileys (e.g. *=)*), abbreviations (e.g. *LOL*), or fashionably misspelled words (such as *'BOOORED!!!!!!!!!!'"* or *"H-A-P-P-Y"*). In other words, it might be that current techniques of automatic interpretation of users' moods are too limited to be reliably used in automated sentiment analysis of Facebook status updates.

A third limitation of FGNH was highlighted by Kramer (2010). He eliminated from his analysis words related to specific occasions. For example, 'happy' in 'happy new year' was not considered to be a genuine reflection of personal happiness but rather the use of a conventional phrase. It seems, however, that a similar approach was not applied in FGNH calculation – possibly to maximize its peaks around "special days" and thus increase its face validity. It is possible that FGNH "holiday" peaks would disappear if Kramer's approach would be implemented.

Fourth, there is a growing body of evidence stemming from social comparison theories suggesting that individuals exposed to positive or negative emotions in their friends' status updates might experience reversed feelings. Consequently, being exposed to a high number of positive status updates published by friends may lead to a general decrease in happiness, while waves of negative status updates may result in people feeling relatively better off (Baumeister & Vohs, 2004; Jordan, et.al, 2010).

Finally, contrary to Kramer's (2010) reasoning that Friday is on the face of it a positive day, it might not be the case that people are happier on holidays and on particular weekdays. Several studies showed that the bluest day of a week is Wednesday and not Monday, while Sunday, rather than Friday, constitutes the peak of the weekly mood cycle (e.g. Charles, 2008).

A limitation of the comparison between FGNH and SWLS is that the two scales tap distinct sources of information on wellbeing; status updates reflect affective responses to recent experiences, whereas SWLS items concentrate on whether an individual believes that their aspirations have been met. These overlap but do not necessarily coincide (Veenhoven, 1984), although short-term changes in SWLS may better reflect affective responses (Schimmack, Diener, & Oishi, 2002; Schwarz & Strack, 2003).

To conclude, we did not find evidence that the FGNH in its current state is a valid measure of national happiness and well-being. This challenges the assumption that linguistic analysis of internet messages is related to underlying psychological states (Golder & Macy, 2011; Miller, 2011). This however does not mean that FGNH is of no use to research. Since hundreds of millions of users generate and read status updates every day, they are likely to affect what people do and feel. Rather than reflect the national mood, status updates could affect the national mood.

References:

Bates, W. (2009). Gross national happiness. *Asian-Pacific Economic Literature, 23*(2), 1-16. doi: DOI 10.1111/j.1467-8411.2009.01235.x

Baumeister, R. F. & Vohs, K. D. (2004). *Handbook of self-regulation : research, theory, and applications*. New York: Guilford Press.

Charles, A. S. (2008). (Tell me why) I don't like Mondays: Does an overvaluation of future discretionary time underlie reported weekly mood cycles? *Cognition & Emotion, 22*(7), 1228-1252.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542-575.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality and Social Psychology, 69*, 71-75.

Facebook Data Team. (2010). Continuing our Study of Happiness, from http://www.facebook.com/notes/facebook-data-team/continuing-our-study-of-happiness/375901788858

Facebook Data Team. (2011). Gross National Happiness, from http://apps.facebook.com/gnh_index/

Gamon, M. (2004). Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th International Conference on Computational Linguistics*, 841 at the 28th international conference on Human factors in computing systems, ACM.

Golder, S.A., & Macy, M.W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, *30,*1878-1881.

Kramer, A. (2010). An Unobtrusive Behavioral Model of "Gross National Happiness". Paper presented at the the 28th international conference on Human factors in computing systems, ACM.

McAllister, F. (2005). Well-being concepts and challenges, *Sustainable Development Research Network Discussion paper*, December, London. Available at: http://www.sd-research.org.uk/wp-content/uploads/finalwellbeingpolicybriefing.pdf.

Miller, G. (2011). Social Scientists Wade Into the Tweet Stream. *Science, 30*, 1814-1815

Mishne, G. & de Rijke, M. (2006). Capturing global mood levels using blog posts. In AAAI 2006 spring *symposium on computational approaches to analysing weblogs (AAAICAAW 2006)*.

Mullen, T. & Malouf, R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *Proceedings of the AAAI symposium on computational approaches to analyzing weblogs*, 159–162.

Nasukawa, T. & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing, *Proceedings of the 2nd international conference on knowledge capture*. Florida, USA, 70–77.

Pang, B. & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the annual meeting on association for computational linguistics* (ACL), 271–278.

Pavot, W., & Diener, E. (2009). Review of the SatisfactionWith Life Scale *Assessing Well-Being: The Collected Works of Ed Diener*. 101-117.

Pennebaker, J. W., Chung, C. K., Ireland, M., A Gonzales, & Booth, R. J. (2007). The development and psychological properties of LIWC2007. 1-22. Retrieved from Linguistic Inquiry and Word Count website:

Pennock, M. (2009). Measuring the progress of communities: Applying the Gross National Happiness Framework. *Measuring the Progress of Societies 2009*, 9-10.

Pennock, M. & Ura, K. (2011). Gross national happiness as a framework for health impact assessment. *Environmental Impact Assessment Review, 31*(1), 61-65. doi: DOI 10.1016/j.eiar.2010.04.003

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-Satisfaction is a Momentary Judgment and a Stable Personality Characteristic: the Use of Chronically Accessible and Stable Sources. *Journal of Personality, 70*(3), 345-384.

Schwarz, N., & Strack, F. (2003). Reports of subjective well-being: judgmental processes and their methodological implications. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: the foundations of hedonic psychology* (pp. 61-84): Russell Sage Foundation Publications.

Stillwell, D. J. & Kosinski, M. (2011). myPersonality Research Wiki. *myPersonality Project.* Retrieved June, 2011, from http://mypersonality.org/wiki

Testing LIWC Online. (2011), from http://liwc.net/liwcresearch07.php

Ura, K. (2008). The GNH Index, from
http://www.grossnationalhappiness.com/gnhIndex/gnhIndexVariables.aspx

Veenhoven, R. (1984). Conditions of Happiness. Dordrecht, The Netherlands: *Reidel*.

Tables:

	Correlation with SWLS
Aggregated by day	
FGNH	-.08
POSITIVITY	-.07
NEGATIVITY	$.13^{*}$
Aggregated by week	
FGNH	-.19
POSITIVITY	-.13
NEGATIVITY	$.37^{**}$
Aggregated by month	
FGNH	$-.64^{*}$
POSITIVITY	-.49
NEGATIVITY	$.72^{**}$

$^{*}p<0.05$; $^{**}p<0.01$

Table 1. The Pearson product-moment correlation between SWLS scores and FGNH aggregated by day, week, and month.

Figure captions:

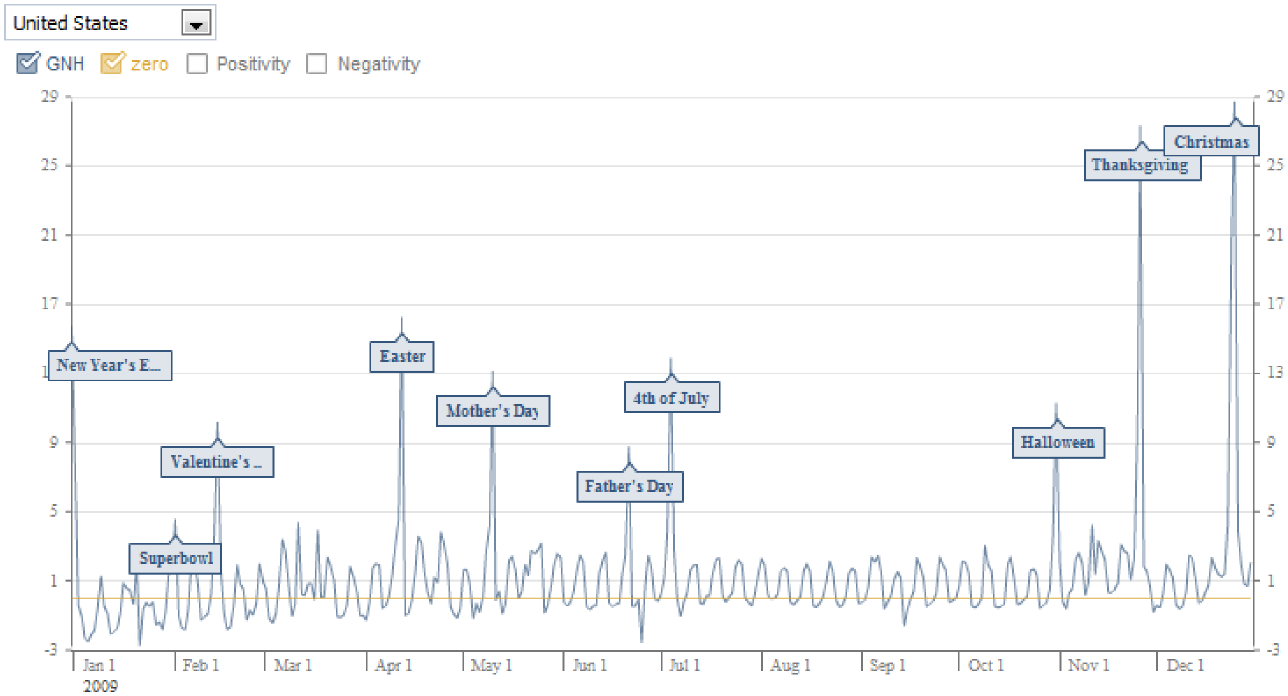

Figure 1. Facebook's Gross National Happiness measure for the United States in 2009 (Facebook Data Team, 2011).

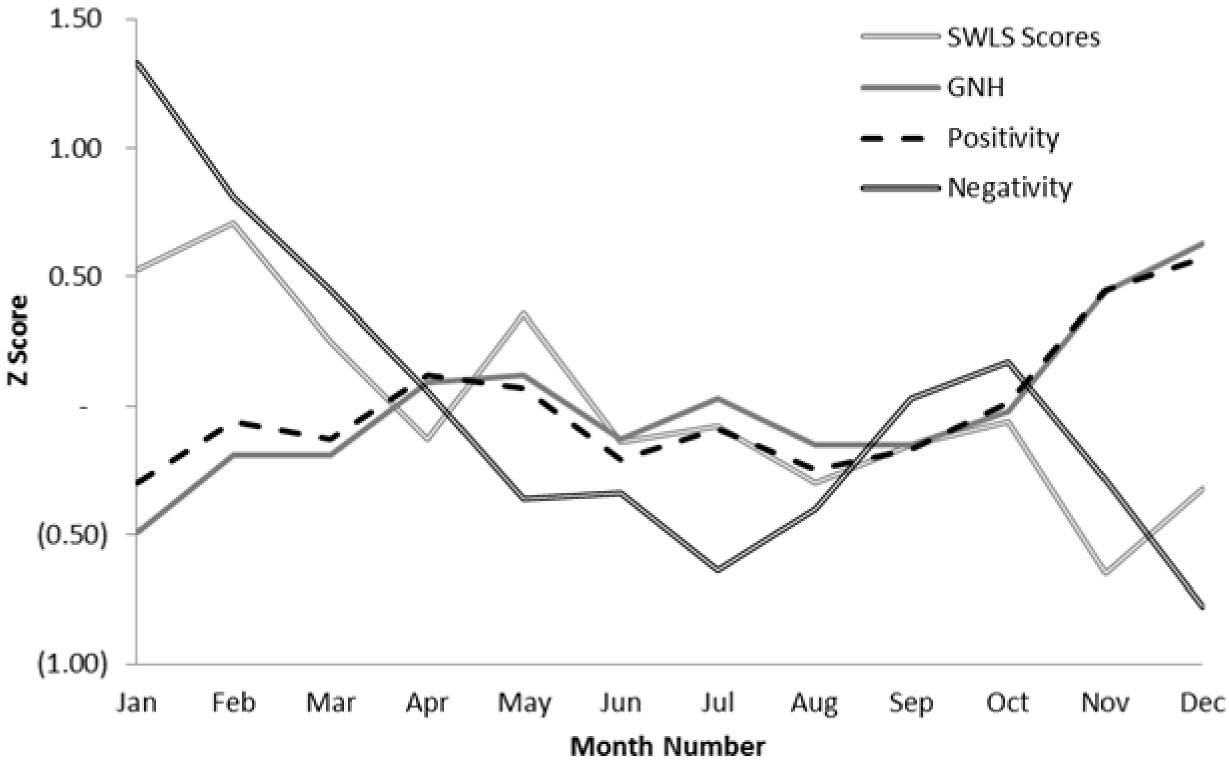

Figure 2. The relation between aggregate SWLS scores, FGNH, Positivity and Negativity by month in 2009.

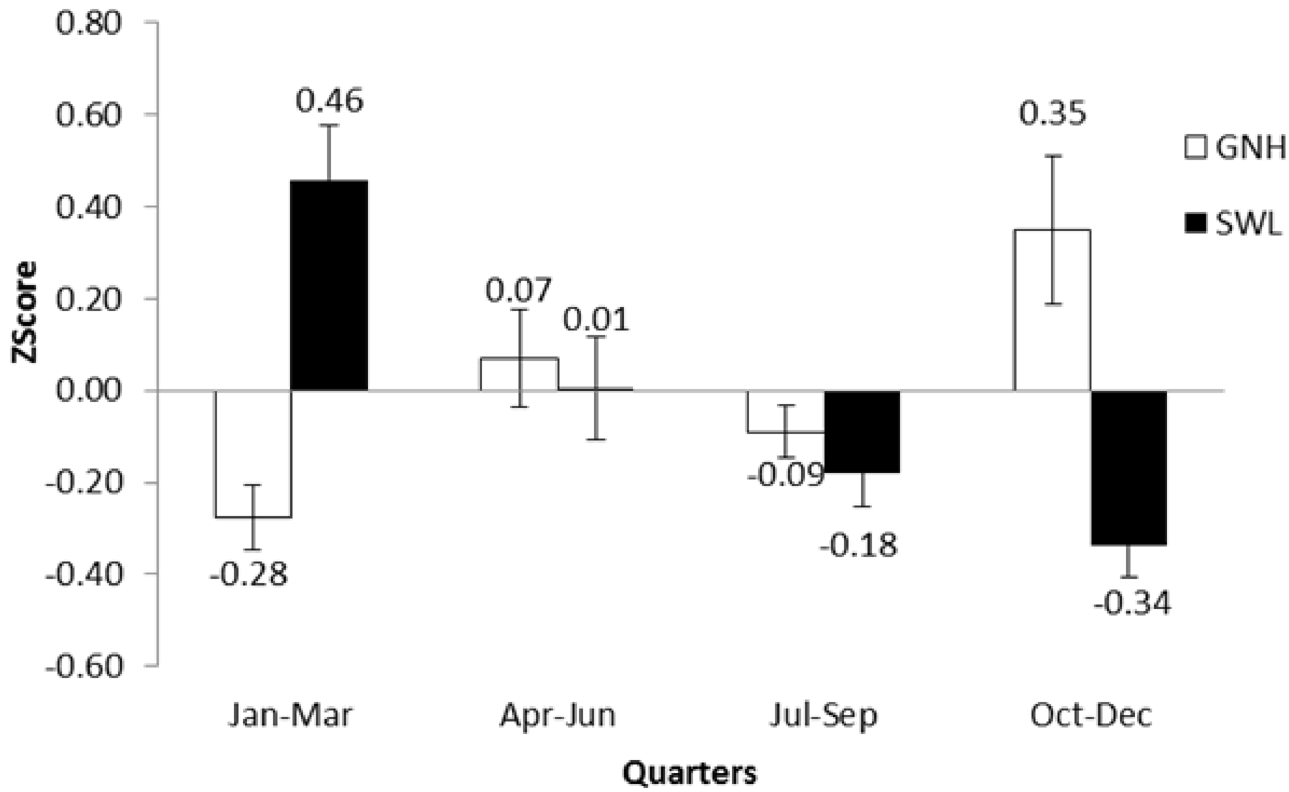

Figure 3 – The means of aggregated SWLS scores and FGNH by quarter in 2009 (error bars: 95% confidence intervals).

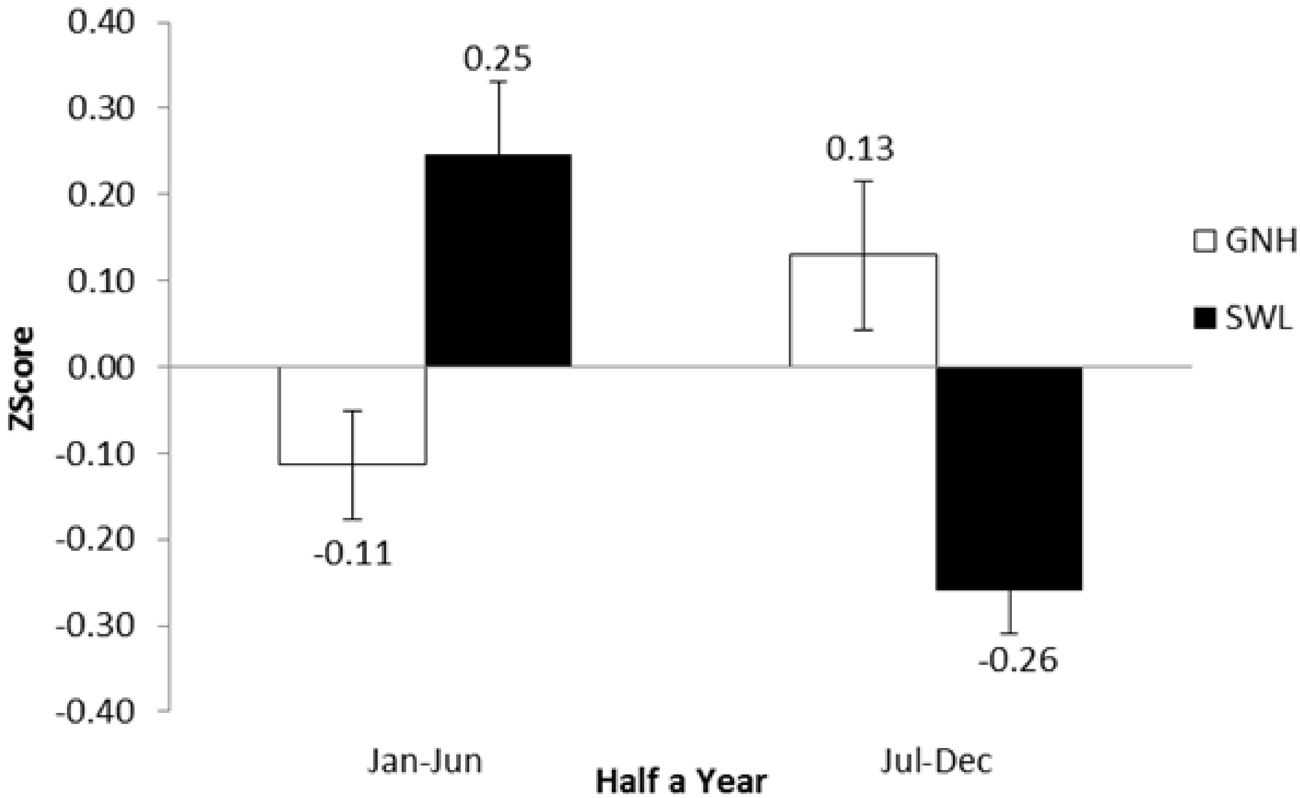

Figure 4 – The means of aggregated SWLS scores and FGNH by half year in 2009 (error bars: 95% confidence intervals).