Institute of Education

UCL

An Investigation of Item Non-Response and Measurement Equivalence in a Mixed-Mode Cohort Study: Findings from the Next Steps Age 25 Sweep

By Lisa Calderwood[1], Alissa Goodman[1], George B. Ploubidis[1], Joseph W. Sakshaug[2], Richard J. Silverwood[1]

[1] Centre for Longitudinal Studies, UCL - Institute of Education, University College London [2] Institute for Employment Research and University of Mannheim

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Contents

Abstract

Mixed-mode data collection is becoming standard practice in survey research. Despite this, there are still unknowns regarding how different modes impact the quality of survey measurements, particularly measurements of multi-item scales. In this study, we investigated the impact of mode on item non-response and measurement equivalence in wave 8 of the Next Steps cohort study, which implemented a sequential mixed-mode (web-telephone-face-to-face) design. After controlling for mode selection via an extensive data-driven variable selection and weighting procedure, we performed a multi-group confirmatory factor analysis to assess measurement equivalence on a well-known scale of psychological distress. The overall findings indicate: 1) a lower proportion of respondents who contribute to item non-response in the web mode than in the telephone and face-to-face modes; and 2) a high-level of measurement equivalence for the multi-item scale across the three modes. Practical implications of these results and research extensions are discussed in conclusion.

Introduction

Using multiple modes to collect data from respondents is a common practice in survey research (De Leeuw 2018). In particular, the practice of deploying multiple modes in sequence is used in several large-scale, policy-relevant longitudinal surveys, including the UK Household Longitudinal Study and the US Health and Retirement Study, as well as the Next Steps (previously known as the Longitudinal Study of Young People in England) cohort study, which is the focus of the present investigation.

Sequential mixed-mode designs have several advantages over unimode designs. First, they can lead to cost savings, particularly if a self-administered mode (e.g. web, mail) is introduced in an otherwise interviewer-administered survey and a significant proportion of people respond in this less-expensive mode (Bianchi et al. 2017; Wagner et al. 2014). Second, offering a second or third mode can lead to higher rates of participation, especially among persons who are reluctant to participate via the first offered mode. And, third, mixed-mode designs may reduce the risk of nonresponse bias by bringing in a more diverse set of respondents relative to a single-mode design (De Leeuw, 2005). All of these advantages have led to a surge in the use of mixed-mode designs and it is likely that such designs will remain a mainstay in survey research for the foreseeable future.

However, despite these advantages, mixed-mode designs have some pitfalls. Specifically, there is a growing body of literature suggesting that the mode of data collection can influence how respondents answer (and don't answer) survey questions. De Leeuw (2005) describes two key dimensions which can lead to respondents providing different answers: the presence vs. absence of an interviewer and the presentation of the questions (oral vs. visual). We briefly discuss each of these dimensions and their potential impact on measurement. Regarding interviewer presence, we know that interviewers have an important effect on social desirability. That is, respondents in interviewer-administered surveys (e.g. face-to-face, telephone) are more likely to answer questions in a more positive light compared to respondents in self-administered (e.g. web, mail) surveys (Kreuter, Presser, and Tourangeau 2008). On the other hand, interviewers help keep respondents engaged and motivated dur-

ing the interview, which may explain why item non-response is less of an issue in interviewer-administered surveys as it is in self-administered surveys (Duffy, Smith, Terhanian, and Bremer 2005; Heerwegh 2009; Jäckle et al., 2015; Lesser et al. 2012).

Regarding the presentation of the survey questions, interviewers typically read the questions and response options out loud to respondents, whereas in self-administered surveys respondents read the questions/response options to themselves (either on a screen in the case of a web survey, or on a paper questionnaire in the case of a mail survey). This dichotomy of visual versus oral presentation may lead to differences in the way respondents process the information. For example, it has been theorized that visual modes produce more primacy effects (i.e. a bias towards selecting the first response options) (Krosnick and Alwin 1987) and oral modes produce more recency effects (i.e. a bias towards selecting the last response options) (Smyth et al. 1987).

Given these mode-specific differences, a natural question to pose is whether measurements collected from respondents in different modes within the same survey are comparable. This question is particularly relevant in the context of multi-item scales, which are assumed to equivalently measure a latent construct of interest regardless of the mode of administration. We investigate the validity of this assumption by studying whether mode impacts item non-response and whether the measurement of multi-item scales is equivalent across different modes used in a mixed-mode design. We use data collected from the Next Steps Wave 8/Age 25 cohort study, which implemented a sequential (web-telephone-face-to-face) mixed-mode design, to address the following research questions:

1) To what extent does survey mode influence respondent prevalence to not answer all survey questions in a nationally-representative cohort study?
2) Do multi-item scales show measurement equivalence across sequentially-administered web, telephone, and face-to-face modes?
3) How does measurement quality compare between the self-administered and interviewer-administered survey modes?

Data Source

Next Steps is a well-known national cohort study which follows a representative sample of people born between 1 September 1989 and 31 August 1990. Cohort members were recruited in schools during their adolescence at age 13/14. The target population consists of young people who were in Year 9 in English state and independent schools and pupil referral units in February 2004. The sample design considered schools as primary sampling units and included an oversampling of deprived schools and minority ethnic groups within schools. The issued sample at baseline comprised approximately 21,000 young people with a total of 15,770 persons interviewed at baseline. An additional minority supplement was added at the Age 17 sweep. From ages 15-20, the issued sample consisted of cohort members who had participated at the previous sweep. For the Age 25 sweep, which is the focus of our study, the issued sample included all cohort members who had ever participated in the study.

In the Age 25 sweep, a sequential mixed-mode design was implemented, starting with a request to complete the survey online, followed by telephone and then face-to-face for the remaining non-respondents. A total of 7,707 (out of 15,531) cohort members participated in one of the three modes (web: 4,797; telephone: 690; face-to-face: 2,220), for an overall response rate of 55.1 percent.

We assessed measurement equivalence across the three modes for a single multi-item scale: the 12-item General Health Questionnaire (GHQ-12). The GHQ-12 is a psychometric screening instrument used to identify common psychiatric conditions (Goldberg and Williams 1988). The twelve-item scale (presented in Table 1) measures general, non-psychotic, as well as minor psychiatric disorders. Each item is rated on a 4-point scale to indicate the severity of a particular symptom of mental ill health.

The GHQ-12 was administered via self-completion in both the web and face-to-face modes and via interviewer-administration in the telephone mode. For the face-to-face interviews, the interviewer handed over their laptop to the respondent where they answered the sensitive item battery privately. For the telephone interviews, the GHQ-12 items were administered by the interviewer without a self-completion option. Given that the visual presentation of the GHQ items was virtually the same in both web and face-to-face modes, we hypothesized that the measurement quality of the items would be more similar between these modes compared to the telephone mode.

Methods

Accounting for Mode Selection

Measurement mode effects can be confounded with selection effects when people have different propensities to respond in a given mode. This is typically the case in observational mixed-mode studies where the mode of interview is mainly dictated by the respondent, rather than the study investigator herself. To account for self-selection into mode we employed a data-driven unit non-response weighting adjustment procedure that utilized observed variables collected in all prior waves of Next Steps. Specifically, a two-stage analytic strategy was used. In the first stage, we conducted a series of seven (one for each of the previous Next Steps waves) within-wave multivariable regressions with mode-specific non-response at Wave 8 as the outcome. All variables whose association reached a statistical significance level of 5% were retained for the second stage. In the second stage, each of the retained variables from the first stage was imputed to produce a complete dataset of predictors. After imputation, the retained variables from each wave entered a series of multivariable regressions predicting mode-specific non-response at Wave 8. Thus, mode-specific non-response at Wave 8 was modelled as a function of predictors from a wave adjusted for all predictors from past waves identified in the first stage. For example, when considering predictors of mode-specific non-response at Wave 8 observed at Wave 4, predictors from Waves 1-3 identified in the first stage were controlled for but predictors from Waves 5-7 identified in the first stage were not included in the model. The predictors which remained statistically significant at the 5% level after controlling for predictors from past waves were then retained.

The retained predictors at stage two were then used to create propensity score adjustment weights for mode-specific unit non-response. The propensity to respond via Web, telephone, and face-to-face was calculated separately for each sample unit. Five propensity score subgroups were generated for each mode-specific outcome using quintiles of the calculated propensity scores. The final adjustment weight was then calculated as the inverse of the average propensity score identified in each subgroup. A total of three adjustment weights were created, one for each survey mode. We applied these weights in all subsequent analyses to control for the confounding effects of mode selection and concentrate on measurement mode effects.

Item Non-Response

To address the first research question, we examined the proportion of respondents in each mode who did not answer at least one of the scale items. Given that non-response rates were very low for the GHQ-12 items, we included items from six additional scales measured in Next Steps. These additional scales measured a variety of topics, including alcohol use, Adult Identity Resolution, Locus of Control, recreational activities, volunteering, and bullying. Across the seven scales, 39 items were included in the analysis. We report two versions of item non-responder rates, one adjusted for the sample design, and another adjusted for the sample design plus the mode-specific unit non-response weighting adjustment described in the previous subsection.

Measurement Equivalence Testing

To test if the measurement of the GHQ-12 multi-item scale was equivalent in the three modes (research question 2) we used multi-group Confirmatory Factor Analysis equivalence testing. Mplus 8.3 software was used to test the three most common forms of measurement equivalence (Meredith 1993): 1) configural equivalence, i.e. the factor structure is the same across the different modes; 2) metric equivalence, i.e. configural equivalence holds and the factor loadings are similar across modes; and 3) scalar equivalence, i.e. metric equivalence holds and the intercepts are the same in all modes. If measurement equivalence holds across modes, then it is possible to compare unstandardized relationships (metric) and/or means (scalar) across the modes.

A simple factor model was used with one latent construct (psychological distress) measured by the GHQ-12. The first loading was set to 1 for identification purposes. For testing configural equivalence we allowed the loadings and intercepts to be estimated freely without restriction. For testing metric equivalence, we restricted the loadings to be equivalent across modes. For scalar equivalence, we additionally restricted the intercepts to be equal across modes. Full Information Maximum Likelihood was used to handle item missing data which we assume are Missing at Random given the model of interest. All analyses account for the complex sample design used in Next Steps.

To assess whether measurement equivalence holds, we applied conventional goodness of fit criteria to assess the fit of each measurement model. The fit criteria include the chi-square test statistic (lower is better), Comparative Fit Index (CFI; higher is better), Tucker-Lewis Index (TLI; higher is better), and the Root Mean Square Error of Approximation (RMSEA; lower is better). We focus on changes in these fit measures when adding the constraints at the different modelling steps.

To answer research question 3, we compare the factor loadings and item intercepts between each pair of modes. As the GHQ-12 was administered via self-completion in both web and face-to-face modes, we expected to find smaller differences in the loadings and intercepts between these two modes relative to paired comparisons involving the telephone mode.

Results

RQ1: Item Non-Responders by Mode

Table 2 shows the percentage of respondents who did not answer at least one of the 39 scale items. The web mode yielded the lowest proportion of item non-responders (5.35 percent) followed by face-to-face (8.56 percent), and telephone (10.68 percent). The difference

between the web and telephone modes and the web and face-to-face modes are both statistically significant. The same conclusions hold after applying the mode-specific non-response weighting adjustment. The non-response adjustment yields a slightly higher proportion of item non-responders in all modes, suggesting that persons with a higher likelihood of unit non-response in a given mode are more likely to be item non-responders in the survey.

RQ2: Measurement Equivalence Across Modes

Next, we assess the fit of each measurement model. The goodness of fit criteria, presented in Table 3, show that a configural model is supported. Based on recommended cut-off values (CFI ≥ 0.90; TLI ≥ 0.95; RMSEA < 0.08) the fit criteria suggest good overall model fit and that the factor structure of the latent variable (psychological distress) is maintained across the three survey modes. The metric model is also supported based on the fact that the CFI and TLI both increase (CFI: from 0.952 to 0.968; TLI: 0.949-0.969) while the RMSEA decreases (from 0.072 to 0.056). Thus, there is indication that the factor loadings are similar across the three survey modes. Lastly, we find that the scalar model – the model with the most constraints – is supported due to the increase in TLI (from 0.969 to 0.974) and decrease in RMSEA (from 0.056 to 0.051). Although the CFI decreases (from 0.968 to 0.967), the change is well within the accepted limit of 0.01. Therefore, we conclude that the factor structure (configural), the factor loadings (metric), and the item intercepts (scalar) are the same in all three survey modes for the GHQ-12 item scale.

RQ3: Comparison of Measurement Quality Between Modes

We now examine the factor loadings and intercepts to assess measurement quality between modes. We hypothesized that the two self-completion modes, namely, web and face-to-face self-interview, would yield more similarities compared to the mode comparisons involving telephone. The standardized factor loadings for each item, shown in Table 4, can be interpreted as the strength of the relationship between the observed item and the latent factor (psychological distress). The overall mean of the loadings in the web mode (1.184) is very similar to the face-to-face mode (1.178), lending support to our hypothesis. Both are notably higher than the mean factor loading in the telephone mode (0.880), suggesting that the self-completion modes yield stronger relationships with the latent factor. Table 4 also shows differences in the item-level loadings for each mode comparison. The overall mean difference is smallest between the self-completion modes at 0.007, whereas the overall mean difference between the telephone and face-to-face modes and the web and telephone modes are each approximately 0.30. Again, this suggests closer correspondence between the self-completion modes with respect to measurement quality relative to the interviewer mode.

Lastly, we examine the intercepts (converted to probabilities) for each GHQ-12 item. The probabilities can be interpreted as the cumulative proportion of responses to each of the three item categories. For example, for the first item “Concentrate on what doing” the proportion of responses to the first category ($1) in the web mode is about 0.06, the proportion of responses to either the first or second category ($2) is 0.85, and the proportion of responses to either of the first three categories ($3) is 0.98. The remaining proportion (0.02) represents the proportion of responses to the fourth category. These probabilities can be used to examine primacy and recency effects. Here we see higher proportions of responses, on average, for the first few categories in the web and face-to-face modes versus the telephone mode, an indication of stronger primacy effects in the self-completion mode, which is consistent with the literature. Furthermore, the intercept probabilities for the telephone mode tend to be lower than the web and face-to-face modes for the third category ($3), indicating a stronger recency effect in the telephone mode which is, again, consistent with the literature on oral versus visual survey modes.

Conclusions

This study investigated the impact of using a sequential mixed-mode (web-telephone-face-to-face) on item non-response and measurement quality in the Next Steps Wave 8/Age 25 cohort study. The study yielded three principal findings. First, we found that the web mode yielded the lowest proportion of respondents who did not answer at least one or more items across seven multi-item scales. Second, we found that the multi-item General Health Questionnaire (GHQ-12) scale achieves a high-level of measurement equivalence across the three modes. All three measurement models: configural, metric, and scalar models were supported by the data, indicating that the factor structure, loadings, and intercepts are comparable between modes. Lastly, we found support for the notion that the two self-completion modes (web and face-to-face self-interview) produced similar measurement quality (in terms of factor loadings and intercepts) relative to the interviewer-administered (telephone) mode. Stronger primacy effects in the two self-completion modes and stronger recency effects in the telephone mode were evident.

These results are reassuring for the Next Steps cohort study and for the use of mixed-mode designs more generally. As longitudinal studies increasingly look to implement mixed-mode designs, it is useful to know that measurement equivalence can be achieved for a well-known scale of sensitive items. Practical reasons (e.g. costs) often dictate the use of mixed-mode designs, particularly those involving online data collection. The fact that adding an online component to an otherwise interviewer-administered survey does not compromise on item non-response nor on measurement equivalence is an advantageous finding from both a practical and methodological perspective.

As with all studies, this one has limitations which could be addressed in future work. For instance, it would be prudent to attempt replication of these results in other mixed-mode studies and on other target populations, including older populations which may not be as web-savvy as the population studied here. Assessing whether measurement equivalence holds for other multi-item scales would also be a useful extension to our work. Lastly, while we adjusted for mode-specific non-response by employing an extensive data-driven variable selection and weighting procedure, it is possible that some unobservable factors influenced respondents' selection into mode. Although we cannot test for this possibility, we nevertheless encourage researchers to make full use of all observable data (which in the case of longitudinal studies, may be immense) in order to make the Missing at Random assumption more plausible.

References

Bianchi, A., Biffignandi, S., and Lynn, P. (2017). Web-Face-to-Face Mixed-Mode Design in a Longitudinal Survey: Effects on Participation Rates, Sample Composition, and Costs. *Journal of Official Statistics*, 33(2), 385-408.

De Leeuw, E.D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21(2), 233–255.

De Leeuw, E.D. (2018). Mixed-Mode: Past, Present, and Future. *Survey Research Methods,* 12(2), 75-89.

Duffy, B., Smith, K., Terhanian, G., and Bremer, J. (2005). Comparing Data from Online and Face-to-face Surveys. *International Journal of Market Research*, 47(6), 615.

Goldberg, D. P. & Williams, P. (1988). *The User's Guide to the GHQ*. NFER-Nelson: Windsor.

Heerwegh, D. (2009). Mode Differences between Face-to-face and Web Surveys: an Experimental Investigation of Data Quality and Social Desirability Effects. *International Journal of Public Opinion Research*, 21(1), 111-121.

Jäckle, A., Lynn, P., and Burton, J. (2015). Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response. *Survey Research Methods*, 9(1), 57-70.

Kreuter, F., Presser, S., and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.

Krosnick, J.A. and Alwin, D.F. (1987). An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly*, 51(2), 201-219.

Lesser, V., Newton, L., and Yang, D. (2012). Comparing Item Nonresponse across Different Delivery Modes in General Population Surveys. *Survey Practice*, 5(2).

Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58(4), 525-543.

Smyth, M.M., Morris, P.E., Levy, P., and Ellis, A.W. (1987). *Cognition in Action*. London: Erlbaum

Wagner, J., Arrieta, J., Guyer, H., and Ofstedal, M.B. (2014). Does Sequence Matter in Multimode Surveys: Results from an Experiment. *Field Methods*, 26(2), 141–155.

Tables

Table 1. General Health Questionnaire (GHQ-12) scale items

Variable	Variable label	Question	Response categories
GHQ_1	Concentrate on what doing	Have you recently been able to concentrate on what you're doing?	1. Better than usual 2. Same as usual 3. Less than usual 4. Much less than usual
GHQ_2	Lost sleep over worry	Have you recently lost much sleep over worry?	1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual
GHQ_3	Playing a useful part in things	Have you recently felt that you are playing a useful part in things?	1. More so than usual 2. Same as usual 3. Less useful than usual 4. Much less useful
GHQ_4	Capable of making decisions	Have you recently felt capable of making decisions about things?	1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less capable
GHQ_5	Constantly under strain	Have you recently felt constantly under strain?	1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual
GHQ_6	Can't overcome difficulties	Have you recently felt you couldn't overcome your difficulties?	1. Not at all 2. No more than usual

			3. Rather more than usual 4. Much more than usual
GHQ_7	Enjoy day to day activities	Have you recently been able to enjoy your normal day to day activities?	1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less than usual
GHQ_8	Face up to problems	Have you recently been able to face up to your problems?	1. More so than usual 2. Same as usual 3. Less able than usual 4. Much less able
GHQ_9	Unhappy or depressed	Have you recently been feeling unhappy or depressed?	1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual
GHQ_10	Losing confidence in self	Have you recently been losing confidence in yourself?	1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual
GHQ_11	Thinking of self as worthless	Have you recently been thinking of yourself as a worthless person?	1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual
GHQ_12	Reasonably happy	Have you recently been feeling reasonably happy, all things considered?	1. More so than usual 2. About the same as usual 3. Less so than usual

4. Much less than usual

Table 2. Percentage of respondents who did not answer at least one of the 39 scale items by mode, before and after applying mode-specific non-response adjustment.

	WEB (1)	TEL (2)	FTF (3)
Before non-response adjustment	5.35[2,3]	10.68[1]	8.56[1]
After non-response adjustment	6.93[2,3]	11.18[1]	10.46[1]

Superscripts denote statistically significant (p<0.05) pairwise differences.

Table 3. Goodness of fit for the multi-item GHQ-12 scale

Model	Chi-square (d.f.)	RMSEA (95% CI)	CFI	TLI	ΔRMSEA	ΔCFI	ΔTLI
Configural	2535.614 (187)	0.072 (0.069-0.074)	0.952	0.949	--	--	--
Metric	1784.261 (209)	0.056 (0.053-0.058)	0.968	0.969	0.016	0.020	0.020
Scalar	1869.912 (255)	0.051 (0.049-0.053)	0.967	0.974	0.005	0.001	0.005

Table 4. Standardized factor loadings and differences between modes

	Factor loadings			Differences between modes		
GHQ-12 items	WEB	TEL	FTF	WEB-TEL	WEB-FTF	FTF-TEL
Concentrate on what doing	1.000	0.597	0.819	0.403	0.181	0.222
Lost sleep over worry	1.030	0.952	0.993	0.078	0.037	0.041
Playing a useful part in things	0.900	0.806	0.785	0.094	0.115	-0.021
Capable of making decisions	0.753	0.514	0.766	0.239	-0.013	0.252
Constantly under strain	0.997	0.911	1.056	0.086	-0.059	0.145
Can't overcome difficulties	1.193	0.726	1.149	0.467	0.044	0.423
Enjoy day to day activities	1.025	1.014	1.127	0.011	-0.102	0.113
Face up to problems	0.797	0.556	0.814	0.241	-0.017	0.258
Unhappy or depressed	1.856	1.552	1.763	0.304	0.093	0.211
Losing confidence in self	1.673	1.252	1.809	0.421	-0.136	0.557
Thinking of self as worthless	1.731	0.753	1.839	0.978	-0.108	1.086
Reasonably happy	1.258	0.925	1.213	0.333	0.045	0.288
Overall mean	*1.184*	*0.880*	*1.178*	*0.305*	*0.007*	*0.298*

Table 5. Intercept probabilities by mode

	Intercept probabilities		
	WEB	TEL	FTF
Concentrate on what doing			
$1	0.06	0.06	0.06
$2	0.85	0.85	0.85
$3	0.98	0.95	1.00
Lost sleep over worry			
$1	0.15	0.15	0.15
$2	0.72	0.66	0.81
$3	0.95	0.91	0.98
Playing a useful part in things			
$1	0.09	0.09	0.09
$2	0.88	0.82	0.88
$3	0.98	0.96	0.99
Capable of making decisions			
$1	0.15	0.15	0.15
$2	0.93	0.94	0.95
$3	0.99	0.99	0.99
Constantly under strain			
$1	0.06	0.06	0.06
$2	0.65	0.58	0.70
$3	0.95	0.92	0.97
Can’t overcome difficulties			
$1	0.17	0.17	0.17
$2	0.84	0.77	0.91
$3	0.98	0.96	0.99
Enjoy day to day activities			
$1	0.03	0.03	0.03
$2	0.84	0.71	0.89
$3	0.99	0.94	0.99
Face up to problems			
$1	0.08	0.08	0.08
$2	0.90	0.88	0.91
$3	0.99	0.99	0.99
Unhappy or depressed			
$1	0.15	0.15	0.15
$2	0.81	0.70	0.90
$3	0.99	0.98	1.00
Losing confidence in self			
$1	0.24	0.24	0.24
$2	0.86	0.81	0.90
$3	1.00	0.97	1.00
Thinking of self as worthless			
$1	0.69	0.69	0.69
$2	0.97	0.91	0.98
$3	1.00	0.98	1.00
Reasonably happy			
$1	0.07	0.07	0.07
$2	0.91	0.87	0.94
$3	0.99	0.99	1.00
Overall mean			
$1	*0.16*	*0.16*	*0.16*
$2	*0.85*	*0.79*	*0.89*

$3	*0.98*	*0.96*	*0.99*