We thank the referee for their detailed review and reply as follows.

Throughout the manuscript the authors imply that variation in dose rate due to variation in aliquot diameter is a novel observation. This isn't true since Spooner and Allsop (2000) investigated… The present paper extends these studies, but the concept of a change in dose rate from the centre to the periphery of a sample holder is not in itself new

The referee makes a good point by highlighting the importance of irradiation field uniformity when geometrical differences of dose rate are detected. We were aware of this issue before starting the experiments and stated this in the text (line 66ff). The sentence may well be insufficient but makes clear that we do not pretend to come up with a new concept on radial dose-rate change. Instead, we present a study which extends the 'size matters' concept (e.g., Duller, 2008) and builds on the conclusion of many studies, ever since single-grain dating is an option: non-uniformity must be considered when calibrating the source (e.g., Ballarini et al.2 006, Veronese et al. 2007). While we hope that this statement cancels the claim of the reviewer, we acknowledge the need for saying more about source uniformity in the text. In sec 2.1 we will describe the sources in terms of (unknown) homogeneity and shape of the radiation field by saying

The ring-shaped sources consist of mini-sources which were tested for homogeneous activity (<5% variation) for the open-ring source before mounting in a circular groove (14 mm diameter) of the stainless-steel housing (Richter et al., 2012). The radiation field of this source varies by 2-8% across 8-10 mm cup diameter (Richter et al., 2012). The larger variation occurs towards the cup edge due to increasing backscatter from the cup rim (Fig. 1), but the inner 6 mm of the cup is exposed to a very homogeneous radiation field (Richter et al., 2012). The sources of the *SMART* readers are not pre-selected for homogeneous activity and may deliver a less uniform radiation field. With a source sample distance of 6.9 mm the radiation field of all sources is expected to be curved. Veronese et al. (2007) show that the dose-rate reduction follows a power function which yields a parabolic curve of variable width. A very wide, hence flat parabolic curve is delivered by the open-ring source (Richter et al., 2012) due to its special design. Before starting the experiments, the beta sources in the readers were manually adjusted to align the centres of sample carrier and source aperture (Discher et al., 2021). Thereby, an almost symmetrical irradiation field across the cup was achieved where the width of the parabola depends on the design of the source.

When considering the experimental data regarding dose rate variation with aliquot diameter, the authors appear to assume that the activity across the face of the source is homogenous. If this isn't the case, and the literature contains a number of studies demonstrating the existence of inhomogenous sources (e.g. Ballarini et al., 2006; Pawlyta et al., 2019), then effects attributed to aliquot diameter or irradiation geometry may actually result from non-uniform distribution of radioisotopes across the face of the source.

We are sorry that the referee came to this impression. We do not make assumptions on the homogeneity of the sources used for the experiments other than that they provide a bell-shaped irradiation field and this is confirmed by the simulation (Fig. 8). Our intention was to show that sources should be calibrated for every irradiation geometry used in dating application. Even if inhomogeneity is quantified using a scintillator or similar tools, it would only confirm the take-home message of our technical note: calibrate every possible irradiation geometry that you use in dating

application because the inhomogeneity itself cannot be addressed by the practitioner. Furthermore, our simulation was conducted on the basis of a homogeneous source, i.e., radioisotopes are evenly distributed across the source face. Details of the simulation results displayed in Figs S3-S6 and summarised in Figs S7 and S8 show that there is a dose enhancement as aliquot size decreases. No doubt, there is more to do to unravel the complex interaction, but we feel we have shown enough data that support our take-home message.

…data were produced using only one example of each source type, meaning that it is impossible to distinguish between source inhomogeneity and other effects.

We feel this statement misses the aim of our study which was clearly outlined in the introduction (unexplained overdispersion of calibration data, etc). Our aim was not to study variations of individual source types. That said, we endorse the statement: our experimental data are not good enough to differentiate between source inhomogeneity and other effects because the 95% CL uncertainties do not allow to disentangle the interplay of various parameters. Here again, the argument the simulation results should satisfy the critique: in the simulation beta particles emitted by the source were evenly distributed – for details see our previous answer.

As a minimum, the authors should acknowledge this potential limitation, though the best solution would be to test the homogeneity of the sources directly (e.g. Pawlyta et al., 2019) or make the same measurements on multiple examples of each source type to determine whether the pattern of change remains constant

Yes indeed, only one example of each source type was used for the experiments and only two source types (planar and closed-ring) were simulated. The objective of the study was to identify the impact of some parameters on the calibration value, we did not intent to study how variable a particular source type is as a result of the manufacturing process. Given that source (in-)homogeneity is an unchangeable fact for the practitioner, Pawlyta's scintillator method remains useful, but does not supersede our calibration approach. Moreover, the scintillator method seems to require direct contact between humans and source (when inserting it in a specially designed socket), hence radiation protection permission and specialised software for data analysis – all together not really readily available in most luminescence laboratory, whereas our recommendation can be implemented straightaway by every practitioner.

Dose rate homogeneity across the face of the sample holder (Spooner and Allsop, 2000) and the variation of dose rate with grain size (e.g. Armitage and Bailey, 2006) are both through to vary with source-to-sample spacing, and possibly also with the dimensions of the active face of the beta source (Spooner and Allsop, 2000).

We think this statement mixes a number of different parameters in a way that is misleading. Dose-rate uniformity across the face of the sample holder does not really exist, at least not for the source-probe distances typically used in luminescence readers for which the radiation field is curved. According to Veronese et al. (2007) this curve takes the form of a parabola. The dose rate changes with changing distance, but for the source-probe distances typically used in luminescence readers

the impact on the shape of radiation field is negligible (Spooner and Allsop, 2000; Veronese et al., 2007). The dimension of the active face is as important as the source aperture is, the design of the source housing and its material. The details are described in the references relevant for each source (e.g. Richter et al. 2012).

It would be very helpful to the reader if this information was given for each instrument and source used in the present study, since it would allow more meaningful interrogation of the data presented. For example, does the "open ring" have a larger active face relative to the size of the sample holder than for other sources? If so, this might explain why aliquot size does not appear to be an important consideration for this source.

Thank you for indicating a weakness in the manuscript. First, for the open-ring source we should have cited Richter et al. 2012 (ATL) - apologies for providing the wrong reference. Second, we should have highlighted the radiation field of this source in comparison to the two other sources. We will add this in sec 2.1 as stated above.

The number of aliquots used to produce each datapoint in the experimental datasets is often rather low. For example, only two 8 mm diameter aliquots of R108_4 were measured using the lexsyg SMART closed ring source….. I accept that the calibration samples used are bright and highly reproducible, but reliance on very small numbers of aliquots makes the resulting dataset prone to other sources of uncertainty e.g. non-uniform aliquot preparation, grain(s) between the sample and hotplate, deformed sample holders (Duller et al., 2000) etc etc. The authors should at least comment on why they think that the small number of aliquots measured doesn't pose a problem.

The number of measured aliquots is indeed small, but, as the referee states correctly, the material is "highly reproducible". If errors are assumed to be independent and normally distributed, the small number of aliquots is not an issue as long as the uncertainties are calculated using the student's t distribution. The values with their uncertainties are then representative of the measured effects and comparable to any other data set. We will re-calculate the uncertainties. It should be said at this point that total uncertainties include the applied gamma-dose uncertainty (see Tribolo et al. 2019 for calculation). However the experimental uncertainties are calculated, for the purpose of our study they are too big to work out the 'size matters' calibration issue, hence the need for simulations.

Errors in Figs and Tables

We thank the reviewer for helping with tidying up the manuscript. We will iron out the issues.

why (and how?) did you perform the GEANT4 simulation without the sample cup?

The purpose was to show the magnitude of electron backscatter without sample. It should have been discussed in the text and we apologies for this oversight. We will include this in sec 4.4 where it says: This [shape of source] is confirmed when simulating charge build-up as a function of depth in aliquot (Fig. 7). Beyond the depth of ca 150 mm the magnitude of the build-up depends on aliquot size and source shape: the increase of dose rate is small in large aliquots irradiated by the closed ring source and significant in medium to large aliquots irradiated by the planar source. It is negligible in small aliquots regardless the beta-source shape. For shallower depths (<150 mm) the magnitude of build-up is enhanced by the electron backscatter of the ss-cup (Fig. 7).

don't really understand the paragraph starting on line 234, which is problematic because it appears to be critical to your explanation of grain size effects with small aliquots. Please expand and clarify.

We will add a figure in the supplement which illustrates the relationship between grain, aliquot and cup with respect to incident and escaping beta particles (Fig. S9). The text (sec 4.5) will say:

Beta particles interact with the aliquot and create secondary electrons that scatter around the interaction point. In the central part of the aliquot the secondary particles interact with neighbouring grains or escape by the surface of the aliquot. If, however, the primary interaction occurs near the aliquot edge, the scattered electrons can, in addition, escape through the edge of the aliquot. The smaller the aliquot, the bigger the percentage of escaping secondary electrons. Furthermore, the thicker the aliquot, the smaller the percentage of secondary electrons escaping by the aliquot surface while the escape pathway via the edge remains the same. The edge effect is therefore governed by the ratio grain size to aliquot size: the bigger the grain and the smaller the aliquot the larger the dose-rate loss. In fact, the simulation shows that the number of scattered electrons decreases for the edge grains (Fig. 9). Thus, the edge effect counteracts the average beta-dose rate increase that occurs for decreasing aliquot sizes because of the radial increase of dose rate towards the centre of the cup, Possibly, it even reverses it when the ratio of grain size to aliquot size is appropriate and the grains are situated sufficiently away from the rim of the cup.

Line 65: The phrase "ring shaped source closed to the top" is awkward. As I understand it there is an "open ring" where the active face is a donut shape constructed of 17 "minisources" whereas the "closed ring" is an approximately circular active face constructed from 23 "mini sources". Diagrams in the SI would help clarify this point and are probably required to answer point 3 above anyway.

We understand the desire to see the design of the sources tested here. For every source a reference is provided where the design is illustrated in great detail. We feel there is no need to copy these published figures into our manuscript. As stated above, the dimension of the source aperture matches the size of the sample carrier.

Line 99: Please explain the rationale for using different stimulation powers for different sized aliquots and show which power was used for which size.

The stimulation power of the blue LEDs (458±5 nm) was reduced with increase of aliquot size to avoid over-exposure of the photomultiplier. A sentence about over-exposure will be added.